Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

THE JOHN CRERAR LIBRARY
JUL 11 1952

CONTENTS

Studies of Distributed Practice: VI. The Influence of Rest-Interval Activity in Serial Learning: B. J. UNDERWOOD 329

Correlation Between Two Psychomotor Tasks as a Function of Distribution of Practice on the First: B. REYNOLDS 341

A Projective Measure of Need for Affiliation: T. E. SHIPLEY, JR. AND J. VEROFF 349

Degree of Conditioning of the GSR as a Function of the Period of Delay: C. T. WHITE AND H. SCHLOSBERG 357

Motivational Properties of Frustration: I. Effect on a Running Response of the Addition of Frustration to the Motivational Complex: A. AMSEL AND J. ROUSSEL 363

Visual Perceptual Latency as a Function of Stimulus Brightness and Contour Shape: P. G. CHEATHAM 369

Performance Decrement in a Simple Motor Task Before and After a Single Rest: E. A. BILODEAU 381

Transfer of Work Inhibition in Motor Learning: GREGORY A. KIMBLE 391

American Psychological Association

Vol. 43 No. 5 May 1952

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., Prince and Lemon Sts., Lancaster, Pa., or 1515 Massachusetts Avenue, N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for behavior theory. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, a single experimental study or an integrated series of studies accomplished simultaneously must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. However, under the new APA reprint policy no free reprints will be supplied to authors submitting manuscripts after January 1, 1952. All of the cost of the author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles must undergo the usual review by the Consulting Editors of the JOURNAL. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

The JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, **41**, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, including tables, footnotes, references, and titles for figures. Tables, figures, footnotes, references, and titles for figures must be placed on separate pages at the end of the manuscript.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 43, NO. 5 MAY, 1952

STUDIES OF DISTRIBUTED PRACTICE: VI. THE INFLUENCE OF REST-INTERVAL ACTIVITY IN SERIAL LEARNING

BENTON J. UNDERWOOD[1]

Northwestern University[2]

In two previous studies (**13, 14**) of serial learning it was observed that overt errors occurred with greater frequency for distributed practice than for massed practice, although learning was faster under the distributed conditions. This finding regarding error frequency was broadly interpreted as conforming to that aspect of Gibson's (**2**) theory which postulates recovery of error (generalization) tendencies with the passage of time. However, no evidence was found to support extensions of the Gibson theory to account for facilitation in learning produced by distributed practice. Indeed, in paired-associate learning and retention of adjectives (**12**) some evidence opposed to the generalization theory was found, e.g., errors did not increase in frequency with retention intervals of as long as 48 hr. Thus, in these previous studies, the role of errors in the learning process remained unspecified; the finding that more errors occurred with distributed than with massed practice of serial lists could not be related satisfactorily to theory accounting for faster learning by distribution.

In discussing this impasse with Dr. C. P. Duncan, it was suggested that the rest-interval activity used in distributed practice might in some way be related to error frequency. With this possibility in mind we proceeded to examine past studies in order to describe the limiting conditions for the greater error frequency under distributed trials. These conditions may be summarized as follows:

1. Serial adjectives with high intra-list similarity; more errors under distribution (as compared with massed) and facilitation in learning by distribution (**14**).
2. Serial adjectives with low intra-list similarity; more errors under distribution with poorly practiced *Ss* only, and a slight facilitation by distribution (**14, 15**).
3. Paired adjectives of both low and high intra-list similarity; no difference in error frequency and no facilitation by distribution during learning (**12**).

In the above experiments color naming has been used to fill the short intertrial rest periods defining distrib-

[1] R. L. Morgan supervised the gathering of the data; E. J. Archer has done most of the statistical work.

[2] This work was done under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research.

uted practice. In our experiments with serial learning these intervals have been no longer than 60 sec.; in paired-associate learning, 120 sec. The colors, small patches arrayed on a board directly above the memory drum, have been named at a rate varying from 60 to 90 per min. as timed by a metronome. Color naming is believed to prevent most rehearsal (although this is probably not important, e.g., **9, 15**), and at the same time provide no interference with learning. It now became our task to examine color naming with the idea of conceptualizing how it might increase error frequency in distributed practice in conformance with the facts listed above. Such a conception was reached which appears to square with these facts.

In color naming *S makes a response to every stimulus;* whether the colors are on a drum or on boards (as in our experiments), *S* responds with the name of the color when the stimulus is perceived. Assume that this develops a set to respond more frequently to other stimuli including words on the drum. If such a set does carry over from the color naming to learning we would expect *S* to respond more frequently than without color naming. This set to respond should produce more errors and it could produce more correct responses. Let us examine this hypothesis in light of the above facts.

In serial learning *S* tries to anticipate the next word whenever a new one appears. He is, in effect, instructed to try to make the correct response to every stimulus but at the same time is told not to be afraid of making errors. A responding set, produced by color naming, should tend to enhance this set produced by instructions. Furthermore, we would expect the errors to be most frequent when the items within the list are least discriminable. In such a situation there are more generalizing tendencies than in a list of high discriminability and these might easily result in overt responses under the instigation of a set to respond produced by color naming. The responding set should, therefore, produce more errors in learning serial lists of nonsense syllables than in learning serial lists of adjectives; more errors should be produced in learning adjective lists of high similarity than those of low similarity. With lists of very low similarity we would expect the responding set induced by color naming to have but little effect except as serial position generalization may occur.

The hypothesis will account for most of the facts available. In our studies of serial adjectives we get a greater difference in error frequency between massed and distributed practice when intra-list similarity is high; the present study will show that errors occur with even greater frequency as a consequence of color naming in serial nonsense-syllable lists. Two other studies of distributed practice in serial learning are available which report frequency of overt errors. Patten (**8**), using extremely well-practiced *S*s, showed no difference between error frequency of massed and distributed learning. However, Patten does not specify the activity used to fill the 2-min. intertrial interval which he used for distribution. Tsao (**11**), with moderately practiced *S*s, showed greater error frequency with distributed trials (30, 60, 120 sec.), but unfortunately he too does not specify the rest-interval activity.

As mentioned above, with paired adjectives we have not found a greater frequency of errors with distributed practice (as compared with massed), even with high intra-list similarity. The reason for this, we believe, is that in paired-associate learning *S* does not respond to every word as it appears: he responds to every other word. Thus, when the stimulus appears he is to try to anticipate the response but when the response appears with the stimulus he is not expected to anticipate anything. It is likely, therefore, that this set counteracts the responding set produced by color naming. Nevertheless, it is conceivable that with paired nonsense lists, where discrimination is low, the

set from naming colors might carry over to the learning task. No data are available on this matter.

Finally, it appears that the set to respond produced by color naming will eventually die out after *S* has had much experience with color naming. In the third study (**13**) in this series serial adjectives of low intra-list similarity were used. The error frequency on the first stage of practice was much greater for the distributed than for the massed conditions, but by the fourth stage (fourth list) the difference had become statistically insignificant. Also, in the first study (**14**), using well-practiced *Ss* with lists of serial adjectives of low similarity, there was no difference in error frequency between massed and spaced conditions. According to our conception, then, we would expect the responding set to decrease as experience in color naming increases.

Confirmation of the hypothesis concerning the effects of color naming would have several implications:

1. If it can be shown that the greater frequency of errors in distributed practice over massed is produced by the color naming per se and not the mere passage of time, there is no need to invoke Gibson's hypothesis (**2**) of recovery of generalization tendencies as we have done previously (**14**).

2. If the hypothesis is confirmed, it is *not* likely that we can infer that facilitation by distribution is produced by color naming only. Other studies (e.g., **10, 16**) of serial learning where color naming has not been used to fill the rest intervals have still shown facilitation by distribution. Yet, it should be clear that according to our hypothesis, any activity which forces *S* to respond continually to all stimuli may produce the same responding set which we are attributing to color naming. It is therefore possible that the gross correlation we have observed between error frequency and rapidity of learning under distributed practice may still maintain itself. That is, while the error increase produced by color naming (or other comparable tasks) may not be necessary for facilitation by distribution, it may heighten the effect.

3. Finally, if any facilitation in learning is produced by color naming it may be important for interpreting certain other phenomena such as reminiscence and warm-up effect in retention. Color naming, or comparable activities, has commonly been used to fill rest intervals (or to serve as warm-up task) in investigating these two phenomena. Furthermore, *Ss* have usually been inexperienced at color naming when reminiscence or warm-up effects were measured.

The purpose of the present investigation, then, is to determine the influence of color naming during the distribution intervals on error frequency and rate of learning. Such an experiment seemed necessary before continuing further systematic studies in this field. For one group of *Ss* we have used color naming to fill the intertrial interval and for a second group we have used symbol cancellation. In symbol cancellation *S* does *not* respond to every stimulus in the series, a situation which should be unfavorable for developing the responding set hypothetically induced by color naming.

Procedure

Lists.—Four serial lists of 14 nonsense syllables each were constructed. The syllables varied from 46.67% to 53.33% association value as determined by Glaze (**3**). The lists were made so that intra-list similarity was as low as possible within the limits imposed by the number of letters available. One list was used for practice, three for experimental conditions. The first item in a list was used as an anticipatory cue only, so that *S* actually learned only 13 syllables. The lists were presented on a Hull-type drum at a 2-sec. rate.

Rest-interval activities.—Two groups of 36 college students each were used. Two independent experiments were performed in which the only difference was the activity filling the inter-trial rests during the distributed-practice

conditions. In one experiment color naming filled the rest intervals (hereafter called Exp. CN), while in the other symbol cancellation was employed (Exp. SC). Each experiment had one practice day, one massed condition, and two distributed conditions.

In Exp. CN *S* named small color patches pasted in rows on a board directly above the aperture of the memory drum. On the practice day the colors were named at the rate of 60 per min. as paced by a metronome, with the speed increased 10 per min. for each successive day so that on the fourth day *S* named at the rate of 90 per min.

In Exp. SC *S* was seated in an arm chair in front of the drum. On the arm were mimeographed sheets of symbols typed in rows. Before each line were three key symbols to be cancelled in that line. On the next line three different key symbols appeared as the ones to be cancelled. The symbols were only 10 in number as follows: ♯, ″, $, %, &, *, ½, @, ¢, ?. Each symbol occurred three times in each row for a total of 30 symbols per row; *S* cancelled (drew a line through) nine of the symbols in each. The order of the symbols varied randomly in each row. In color naming *S* responded positively to each stimulus in order by saying the color; in symbol cancellation *S* responded positively to only 9 of the 30 stimuli. Both activities could be started and stopped readily, as is necessary for the short rest intervals. Brief instructions for these activities were given on the practice day.

Conditions of distribution.—Massed practice consisted of 2 sec. between each trial, being represented by a single blank space on the tape. One distributed condition used a 30-sec. interval between each trial and the other had a 60-sec. interval. It was during these intervals that the rest activity varied for the two experiments. In each condition the list was presented until *S* achieved one perfect recitation of the list. *S* returned after 24 hr. to recall and relearn the list under massed practice.

On the practice day in both experiments the list was learned by massed practice to a criterion of 7 out of 13 items correct on a single trial. Upon reaching this criterion *S* was read instructions for performing the rest-interval activity following which he continued to learn the list to one perfect trial with 30 sec. between each trial. In Exp. SC, of course, the 30-sec. intervals were filled with symbol cancellation, while in Exp. CN they were filled with color naming.

Design.—Each *S* in a given experiment served in all conditions. With three experimental conditions six *S*s are necessary to effect complete counterbalancing. Thus, the 36 *S*s constituted six blocks of six *S*s each. Furthermore, a different order of lists was used for each block so that the lists were also completely counterbalanced. Therefore, any differences in difficulty of the lists could be determined since they would not be confounded with practice effects. Similarly, practice effects could be determined and still be equally distributed over all conditions.

Two additional factors were varied in these experiments. It has been common practice in experiments where *S* serves in all conditions never to have him learn a new list during a session where recall and relearning of another list takes place. Rather, *S* relearns the one list and then returns the following day to learn the new list. Since relearning of a list usually takes place quite rapidly, it would be much more economical to learn the new list after the relearning of the old. This would require fewer sessions for *S* and would make more efficient use of *E*'s time for a given session. It was felt worthwhile to determine whether or not this procedure would significantly influence learning and retention. Therefore, half the *S*s in each experiment always learned a new list during the same session that they recalled and relearned a previous one, while the other half waited 24 hr. before learning the new list. Our analysis shows that these two procedures produced no appreciable differences in learning, recall, or relearning. However, since these data are irrelevant to the main purpose of the experiment, they will not be reported.

Another factor investigated, the results of which will not be detailed, concerns the effect of warm-up on recall. Half the *S*s was given 1 min. of warm-up just before recall, the other half was not. This variable was balanced equally with the others, i.e., half the *S*s in each experiment who had relearning and new learning on the same day had warm-up and half did not. The warm-up task was 1 min. of number guessing as used by Irion and Wham (7). This variable also proved to be an insignificant source of variance. Such a finding does not necessarily contradict previous results (6, 7) since in no other study has warm-up before recall been investigated when *S* serves in several conditions.

To summarize the main conditions: 72 *S*s learned serial lists of nonsense syllables under one massed and two distributed conditions, recalling and relearning each after 24 hr. For 36 *S*s color naming was the rest-interval activity, for the other 36, symbol cancellation. It was our hypothesis that the color naming induces a set to respond to stimuli in the list, hence, should produce more errors than should symbol cancellation which should not develop such a set.

Results

Practice day.—Presenting performance scores for the practice day serves two purposes: (*a*) they will show that the two groups did not differ significantly before the variable was introduced, and (*b*) they will allow initial evaluation of differences produced by rest-interval activities.

Both groups learned the same practice list under identical conditions until seven responses were given correctly on a single trial. The mean number of trials to learn to this criterion and the mean number of errors per trial are given in the left portion of Table 1. It is apparent that the two groups can be considered equivalent; hence, differences which occur subsequently must be a function of differences in rest-interval activity.[3]

The mean number of trials required to learn from the criterion of seven correct to one perfect trial, together with the error analysis, appears in the right portion of Table 1. In terms of trials to learn, the t of 2.40 is significant at about the 2% level of confidence with 70 *df*. It is reasonable to conclude that the introduction of different rest-interval activities produced a difference in rate of learning. An evaluation of the learning curves (Fig. 1) for the practice day does not give a conclusive answer as to whether the differences are a result of facilitation produced by color naming or inhibition produced by symbol cancellation. The inflection seems most prominent in the color-naming curve, but since we have no "true" learning curve with which to compare these, it is difficult to arrive at a firm conclusion. In any event, it will be seen that the difference in learning produced by the rest-interval activities is a very transient affair.

With regard to overt errors (*S* actually saying an incorrect response) in learning from 7 out of 13 to one perfect, it can be seen that the difference is not significant although more errors occur with color naming than with symbol cancellation. However, the error data for the experimental day strongly support the idea that a responding set is being built up during the later part of the practice-day learning.

[3] Product-moment correlations between trials to learn to 7 out of 13 correct on the practice day and total trials to learn under all experimental conditions are .62 for Exp. SC and .34 for Exp. CN. Correlations for total trials on practice day and total trials on experimental day are .60 for Exp. SC and .52 for Exp. CN.

TABLE 1

Evaluation of Practice-Day Performance in Terms of Mean Error Per Trial and Trials to Learn

	To Seven Correct				From Seven Correct to One Perfect			
	Trials		Errors		Trials		Errors	
	M	σ_M	M	σ_M	M	σ_M	M	σ_M
Exp. SC	13.42	.86	1.32	.13	20.06	1.40	1.15	.11
Exp. CN	14.08	1.02	1.38	.11	15.83	1.06	1.41	.10
Difference	.66		.06		4.23		.26	
$\sigma_{diff.}$	1.33		.17		1.76		.15	
t	.50		.35		2.40		1.73	

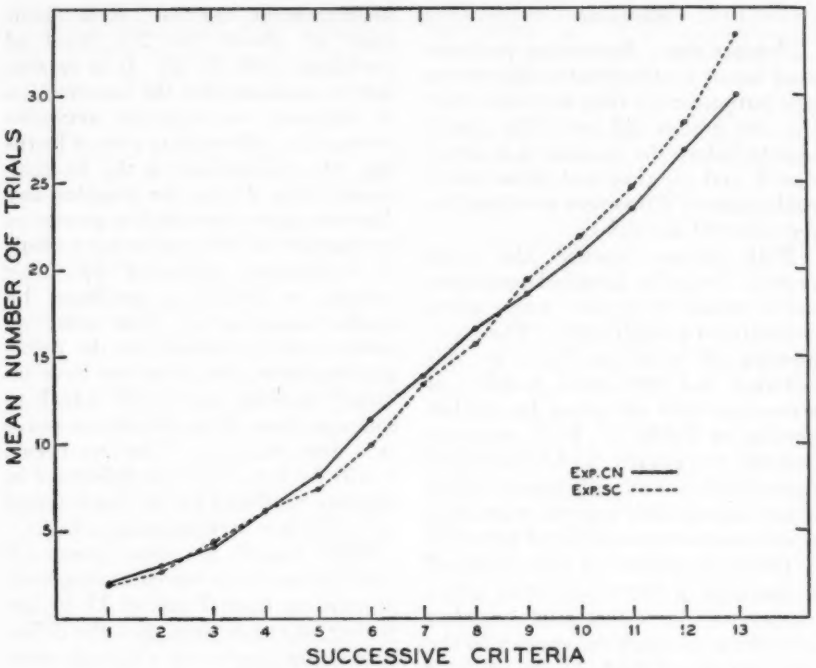

FIG. 1. Learning curves for the practice list

Errors during learning of the experimental lists.—It was shown that after the introduction of the two rest-interval activities on the practice day more errors occurred with color naming than with symbol cancellation. While this difference did not meet reliability requirements, the data for the experimental sessions indicate that it is a real phenomenon. The mean number of overt errors per trial for each condition for the two experiments has been plotted in Fig. 2. It is apparent that many more errors are produced by color naming than by symbol cancellation. The F is 16.24, where an F of 7.01 is needed for significance at the 1% level of confidence with 1 and 70 *df*. Also, the frequency of errors increases as a function of the length of rest. For Exp. CN, F is 8.30, and for Exp. SC, F is 5.53. With 2 and 66 *df* an F of 4.95 is needed for significance at the 1% level. (See Table 2 for the method by which the complete statistical analyses were made for each response measure.)

These differences as a function of intertrial rest intervals must be interpreted cautiously. Looking first at Exp. CN, it can be seen that the error frequency for the 30-sec. and 60-sec. conditions is approximately equal. The significant F is largely a resultant of the low frequency of errors for the massed (2-sec.) condition. But it should also be noticed that the error frequency is greater for the massed condition in Exp. CN than for the

comparable condition in Exp. SC. On the massed conditions, of course, there was no formal rest-interval activity. These facts indicate that the responding set, for which Fig. 2 appears to be strong confirmation, carries over from day to day. Thus, the massed condition would be influenced by previous conditions of distribution in which color-naming was used. At least two further tests of this proposition are available and will be mentioned at this time.

If the responding set does carry over from a distributed to a massed condition, it should be clearly shown by tabulating errors separately for different stages of practice. There are 12 *Ss* who had the massed condition on the first experimental day (preceded only by the color naming in learning the later part of the practice list). Another group of 12 *Ss* had the massed condition on the second experimental day (preceded by one distributed condition), and the remaining 12 *Ss* had the massed condition on the last day, this being preceded by two distributed conditions. If the responding set carries over we should expect the greatest error frequency for the massed condition to occur when that condition is preceded by two distributed conditions, and least when preceded only by the small amount of color naming on the practice day. Since these three subgroups of 12 *Ss* each were not equal in error frequency on the practice day, it is necessary to adjust for these initial differences. Subtracting the mean error frequency for each group on the practice list from the mean error frequency on the massed experimental list gives −.22, .00, and .11 for the groups with massed practice on the first, second, and third experimental days, respectively. The corresponding values for Exp. SC are −.21, −.27, and −.20. Thus, for Exp. CN, the greater the number of distributed conditions before a massed condition the greater the error frequency on that massed condition. The responding set clearly carries over from day to day. No such phenonenon is observed when symbol cancellation is used to fill the interval.

A second finding, to be presented in detail later, also confirms the idea that the responding set transfers from day to day. During relearning the error frequency is consistently greater for Exp. CN than for Exp. SC. Since relearning in both experiments was under massed practice, the greater error frequency with Exp. CN must be a function of transfer of the responding set produced by color naming during the learning under the distributed conditions.

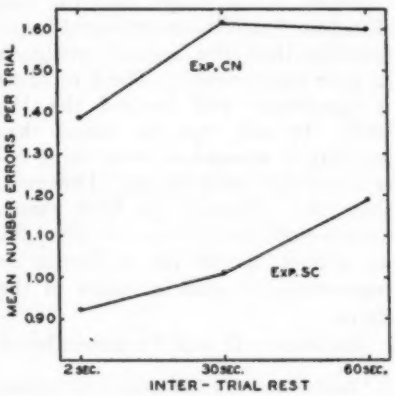

FIG. 2. Errors made in learning to one perfect trial under the three experimental conditions

The next problem concerns the relationship between error frequency and distribution interval in Exp. SC. In Fig. 2 it is seen that error frequency increases with the distribution interval, and it has been shown that this trend is significant. This would seem to substantiate Gibson's theory (2) that generalization tendencies recover with the passage of time even when the effect of color naming is removed. However, in three subsequent experiments we have not found this increase in errors using symbol cancellation to fill the intervals. One of these experiments used the same lists as in Exp. SC, the other two employed lists of considerably higher intra-list similarity. In none of these three experiments does error frequency increase consistently and significantly as a function of the length of interval. We find no adequate accounting for these contradictory findings, but the odds appear to be three to one against accepting the present findings as representing a true phenomenon.

In the introduction it was suggested

that errors as a function of color naming will decrease as a function of experience at color naming. The present data for Exp. CN show that this is confirmed although the amount of color-naming experience given *S* is not enough to give a definite conclusion as to whether or not the error frequency would eventually reach the level which would be observed without color naming. For Exp. CN the mean number of errors per trial (all conditions combined) for the three stages of practice is 1.62, 1.51, and 1.47. The *F* (2.96) falls just short of the value needed for the 5% level (3.14). For Exp. SC the corresponding values are .99, 1.07, and 1.04.

Effect of rest-interval activity on learning.—The mean number of trials to reach one perfect recitation for each condition of each experiment is plotted in Fig. 3. A complete statistical analysis is presented in Table 2. In Table 2 variance due to activity (color naming vs. symbol cancellation), practice effects, lists, intertrial rests (2, 30, and 60 sec.), and *S*s, along with various interactions, are listed separately. In testing for effects of activity, the S*s*

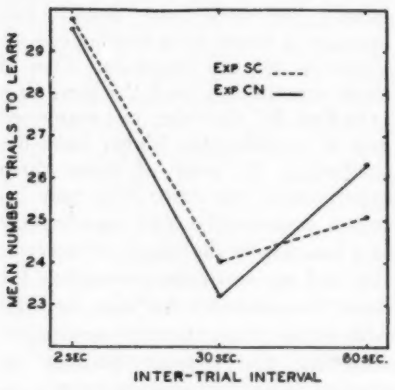

FIG. 3. Trials to learn to one perfect under the three experimental conditions

TABLE 2

ANALYSIS OF VARIANCE OF TRIALS TO LEARN TO ONE PERFECT ON EXPERIMENTAL DAYS

Source	*df*	Mean Square	*F*
Activity	1	10.2269	....
*S*s within activity	70	228.1586	
Intertrial rest	2	730.0603	19.61
Practice	2	923.6991	24.81
Lists	2	93.8519	2.52
Practice × activity	2	59.1713	1.59
Intertrial rest × activity	2	5.8935	....
Lists × activity	2	3.1852	....
Pooled *S*s × practice (residual)	132	37.2294	
	215		

Within Activity is the appropriate error term. For all other tests, *Pooled* S*s* × *Practice* (Residual) is proper. In these latter cases an *F* of 4.78 is significant at the 1% level for 2 and 132 *df*.[4]

From Fig. 3 and Table 2 it can be seen that while rate of learning in Exp. CN is slightly faster than for Exp. SC, the difference falls far short of significance. It may be concluded, therefore, that the two rest-interval activities used here do not differentially influence the learning as observed on the experimental days.

Figure 3 also shows that the two distributed conditions produced faster learning than the massed condition in both experiments. The *F* of 15.69 is significant well beyond the 1% level. It will also be noted that learning is somewhat faster with 30-sec. rest than with 60 sec. However, since the difference for both experiments combined gives a *t* of only 1.90, we cannot accept the difference as representing a true inversion in the curve.

Retention.—It will be remembered

[4] Bartlett's test for homogeneity of variance has been made in all cases where the variance differences appeared large. In no case were the variances significantly different.

that each of the experimental lists was recalled and relearned 24 hr. after original learning. In terms of the performance measures there are no differences produced by the major variables which reach acceptable levels of statistical significance for either recall or relearning. However, as shown in Table 3, two distinct trends are apparent: (*a*) retention is better in Exp. CN than in Exp. SC, and (*b*) retention is better following massed learning than following distributed learning. This second trend is probably indicative of a true phenomenon since in other experiments to be reported later the differences have been considerably enhanced by increase in intra-list similarity. Nevertheless, on the basis of strict statistical evaluation of the present results we must conclude that whether symbol cancellation or color naming is given during original learning is of small consequence for retention, and that with lists of low intra-list similarity, retention does not vary as a function of intertrial interval.

A subsidiary finding of the retention analysis is that recall is inversely related to stage of practice. For Exp. CN the mean number of items recalled at the first stage of practice (recall of the first experimental list after 24 hr.) was 5.28. For the second stage the mean was 4.56, and for the third stage, 3.72. The corresponding values for Exp. CN were 4.50, 3.61, and 3.31. The F (5.07) is significant well beyond the 1% level. Thus, these results confirm previous findings with paired-adjective lists (4).

Concerning error frequency during relearning, the data show consistently more errors for Exp. CN than for Exp. SC. In Exp. CN the total errors at recall were 73, 69, and 67 for intertrial rests during learning of 2, 30, and 60 sec., respectively. For Exp. SC the values were 64, 56, 66. During relearning for all conditions combined for Exp. CN the mean errors per trial was 1.19, and for Exp. SC, .97. The F (4.91) with 1 and 70 df is significant at approximately the 4% level. Thus, there is further evidence that the color-naming set transfers from one day to the next.

TABLE 3

MEAN NUMBER OF ITEMS RECALLED AND TRIALS REQUIRED TO RELEARN

Intertrial Rest	Recall*		Relearning**	
	Exp. CN	Exp. SC	Exp. CN	Exp. SC
2 sec.	5.11	4.03	8.50	9.64
30 sec.	4.08	3.81	9.39	10.11
60 sec.	4.36	3.58	10.58	11.36

* The estimate of σ_M based on residual variance is .42.

** The estimate of σ_M based on residual variance is .74.

DISCUSSION

The two major findings of the present experiments are: (*a*) a consistent and large difference in error frequency for the two different rest-interval activities (symbol cancellation and color naming), and (*b*) no major differences in rate of learning in the body of the experiment as a function of rest-interval activity.

The first finding has been taken as confirmation of the hypothesis that color naming during intertrial rests develops a set to respond which carries over to the learning activity. This results in more errors being made. However, since there is no known "true" error frequency to which we can refer the findings, one may just as legitimately suggest that the differences in error frequency may be due to a depressant effect produced by symbol cancellation. Or, one could postulate that color naming increases errors somewhat and symbol cancellation decreases them somewhat. The subanalyses of the data, however, support most strongly the hypothesis that the differences are due largely to color naming. It has been shown that in Exp. CN there is a carry over of a set from distributed

to massed conditions, and that error frequency decreases as a function of stage of practice. The error frequencies for Exp. SC gave no evidence for such trends. With symbol cancellation the error frequency did not change from day to day and there were no data which showed that a distributed condition (where symbols were cancelled) influenced a massed condition coming later (where symbols were not cancelled). In short, even though the rate of response to stimuli in symbol cancellation was about one in three (*S* cancelled 9 out of 30 symbols) there is no evidence that this developed a set *not* to respond in learning the list. Nevertheless, arguments such as the above cannot be conclusive since by assuming certain "true" error frequencies a defense can be established for an increase in errors because of color naming, a decrease because of symbol cancellation, or a combination of both. We are inclined to accept the first alternative.

The second major finding was that learning did not vary appreciably as a function of the rest-interval activities used. Minor exceptions were noted and these may be of considerable importance. On the practice day the two groups learned at approximately the same rate to 7 correct items out of the 13. At that point color naming was introduced as a rest-interval activity for one group and symbol cancellation for the other. A significant difference occurred in learning to one perfect recitation (Fig. 1 and Table 1). We have interpreted this as being facilitation produced by color naming but again we have no conclusive evidence against interpreting it to be a function of an inhibition or depressant resulting from symbol cancellation. It was also observed that the error frequency for color naming became greater than the error frequency for symbol cancellation after the different rest-interval activities were introduced. And, while this difference was not significant statistically, the large differences in error frequencies on the experimental day justify the interpretation of the difference on the practice day as representing a true phenomenon. Thus, on the practice day we observe small differences in error frequencies and a significant difference in learning. On the experimental days there is a large difference in error frequencies and little difference in rate of learning.

These facts raise the basic issue of the relation of errors to learning. It is apparent that in our gross scores on the experimental days no relationship is present. The product-moment correlations between mean errors per trial and trials to learn for all conditions combined is $-.02$ for Exp. SC and .07 for Exp. CN. Nevertheless, one can defend the proposition that on the practice day there is a gross relationship between error frequency and rate of learning. One might assume that for each *S* there is an optimum number of errors, optimum in the sense that it will lead to most rapid learning. Assume further that the average *S* is working a little below the optimal level when he first starts to learn. Color naming might then initially increase the errors somewhat and lead to more rapid learning. With continued color naming (as on the experimental days) the average *S* goes beyond the optimum error frequency and learning is retarded somewhat, becoming now roughly comparable to that produced by symbol cancellation.

The above hypothesis appears reasonable in view of the relative consistency of an individual in producing errors. For example, the correlation between error frequency on the initial and later part of the practice day for Exp. CN is .65, and for Exp. SC, .68. The correlation between errors on the later part of the practice day and all experimental days combined is .53 for Exp. CN, and .59 for Exp. SC. Color naming, then, if it increases error frequency as we think it does, increases it with some uniformity for all *S*s; the entire distribution simply moves up. It is therefore reasonable to assume that performance in learning is first enhanced and then retarded by color naming because a small increase in errors raises *S* to an optimum error frequency whereas a continued increase begins to interfere with learning. Thus, on the experimental days, both groups take about the same number of trials to learn.

If we are correct in our interpretation that when color naming is a fairly new task for *S* his learning may be enhanced, other phenomena may be brought under such an interpretation. More specifically, reminiscence as usually studied in verbal learning may be in part a function of the responding set produced by color naming. Several modern studies (1) have used color naming to fill the short intervals (30 sec. to 2 min.) which produce reminiscence. Under the responding set produced by color naming near-threshold responses should become overt. This should produce more errors but also more right responses than would occur without color naming. On recall, in the present experiments, where we presume the responding set to be diminished, more errors and

more correct responses were observed in Exp. CN than in Exp. SC. While these differences were not significant, they occurred consistently for all conditions. It is, then, a distinct possibility that some reminiscence in some studies at least may be a product of color naming or comparable activities. Obviously, research directed specifically at this issue will be needed.

The retention data have shown that forgetting is less following massed than following distributed learning. While these differences were not significant it is well to point out that Hovland (5), using comparable materials, found better retention following distributed learning than following massed. However, Hovland used a 2-min. intertrial interval. There is, therefore, no basic contradiction between the two studies but the need for further variation in the length of the rest intervals becomes apparent. The common generalization that retention is better following distributed than following massed learning clearly has limitations. More extensive findings on this matter will be presented in a later report.

Summary

Certain studies have shown that in experiments on distributed practice more errors occur under the distributed conditions than under the massed even though learning is more rapid by distribution. It was hypothesized that the difference in error frequency is due to the color-naming task used to fill the short intertrial rests during distribution. More specifically it was suggested that the color naming induces a set to respond to stimuli more frequently than would be the case without color naming, thus resulting in more errors. In the present experiments one group of 36 *S*s went through one massed (2 sec. between trials) and two distributed (30 sec. and 60 sec. between trials) conditions with color naming as the rest-interval activity. A comparable group went through the same conditions but symbol cancellation was used as the rest-interval activity. The learning material was serial lists of nonsense syllables of low intra-list similarity. Both groups were given one practice list. This list was learned to 7 out of 13 correct on a single trial by massed practice and then by distributed practice (30 sec. between trials) to one perfect recitation. On the experimental days learning was carried to one perfect trial, with recall and relearning by massed practice after 24 hr.

The essential findings were:

1. After the introduction of the two different rest-interval activities on the practice day, a significant difference occurred in rate of learning. More errors appeared with color naming than with symbol cancellation although the difference was not significant.

2. On the experimental days many more errors occurred with color naming than with symbol cancellation on all conditions. However, no difference in rate of learning occurred as a function of the activity. In both experiments learning was more rapid under distributed than under massed practice.

3. The responding set induced by color naming was found to carry over from day to day.

4. Error frequency was found to be greater with the distributed conditions than with the massed.

5. No highly significant differences occurred during recall and relearning as far as learning performance was concerned, but two distinct trends were apparent, namely, better recall

and faster relearning for the color-naming group and better retention following massed than following distributed practice.

6. More errors appeared at recall following color naming than following symbol cancellation and the same difference was observed during relearning.

The results were taken as confirmation of the hypothesis that color naming establishes a responding set for learning which is reflected in an increase in error frequency. Subsidiary evidence suggests that the differences in learning and in error frequency are a function of a positive effect of color naming and not a function of a depressant effect produced by symbol cancellation. It was also suggested that color naming may temporarily facilitate learning performance; if such is the case certain findings concerning reminiscence may be a function of this activity.

(Manuscript received August 30, 1951)

References

1. Buxton, C. E. The status of research in reminiscence. *Psychol. Bull.*, 1943, **40**, 313–340.
2. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
4. Greenberg, R., & Underwood, B. J. Retention as a function of stage of practice. *J. exp. Psychol.*, 1950, **40**, 452–457.
5. Hovland, C. I. Experimental studies in rote learning theory: I. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.
6. Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55**, 336–341.
7. Irion, A. L., & Wham, D. S. Recovery from retention loss as a function of amount of pre-recall warming-up. *J. exp. Psychol.*, 1951, **41**, 242–246.
8. Patten, E. F. The influence of distribution of repetitions on certain rote learning phenomena. *J. Psychol.*, 1938, **5**, 359–374.
9. Rohrer, J. H. Factors influencing the occurrence of reminiscence: attempted formal rehearsal during the interpolated period. *J. exp. Psychol.*, 1949, **39**, 484–491.
10. Tsao, J. C. Studies in spaced and massed learning. II. Meaningfulness of material and distribution of practice. *Quart. J. exp. Psychol.*, 1948, **1**, 79–84.
11. Tsao, J. C. An analysis of the achievements in spaced and massed practice: a study of the inhibition theory. *J. gen. Psychol.*, 1951, **44**, 188–198.
12. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.
13. Underwood, B. J. Studies of distributed practice: III. The influence of stage of practice in serial learning. *J. exp. Psychol.*, 1951, **42**, 291–295.
14. Underwood, B. J., & Goad, D. Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, 1951, **42**, 125–134.
15. Withey, S., Buxton, C. E., & Elkin, A. Control of rest interval activities in experiments on reminiscence in serial verbal learning. *J. exp. Psychol.*, 1949, **39**, 173–176.
16. Wright, S. T., & Taylor, D. W. Distributed practice in verbal learning and the maturation hypothesis. *J. exp. Psychol.*, 1949, **39**, 527–531.

CORRELATION BETWEEN TWO PSYCHOMOTOR TASKS AS A FUNCTION OF DISTRIBUTION OF PRACTICE ON THE FIRST

BRADLEY REYNOLDS[1]

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

It is well known that spacing of training trials is related to level of performance. It is not known to what extent the correlations between performance measures taken at various points in the learning of one skill and performance measures in another skill are affected by the distribution of practice in the training of the first skill. The existence and nature of such relationships have relevance for at least three problems in the field of motor skills.

The first problem has to do with the generalization of the effects produced by massing training trials. If it is assumed that massing training trials accentuates the growth of some inhibitory state, the question which arises is whether such a state persists to some degree following training. If such is the case, will the effect of such a residual inhibitory state be manifested in the exercise of skills other than the specific skill which has been practiced?

A second problem concerns the relationship between performance at one stage of learning with performance at a later stage of learning. It has been observed (**1**; **2**, pp. 1022–1023; **3**; **5**) that correlations between performance scores for earlier periods of practice and those for later periods decrease as a function of the amount of intervening practice. Since inhibition is assumed to be a function of amount of practice, it is possible that such decreased correlation is attributable in part to the development of inhibition. If this is so, such decreases in correlation will be more marked if training is carried out with massed rather than distributed practice.

A third problem has to do with the relationship between the distribution of test trials and the validity and reliability of psychomotor tests. If it is assumed that growth of inhibition contributes to decreased correlation between earlier and later stage performance scores, then it is quite possible that the correlation between scores for two trials occurring early in training will be lower than the correlation between scores for two trials occurring later in training. By way of example, let us consider a ten trial test. It can be assumed that inhibition will have developed by Trial 10 to a higher level than that present at Trial 1. There will be individual differences in the rate of development of inhibition. These will contribute to between-*Ss* differences during Trial 10 to a greater degree than during Trial 1. If the amount of inhibition present at Trial 9 is roughly equivalent to that present at Trial 10, this will contribute to the correlation between Trial 9 and Trial 10. If the amount of inhibition is appreciably greater at Trial 2 than at Trial 1, this will reduce the correlation between Trial

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Air Force.

1 and 2. Thus, the correlation between Trial 9 and Trial 10 will be higher than the correlation between Trial 1 and Trial 2. This has been observed (3). Since inhibition develops as a function of spacing of trials, then with massed trials the correlation between two adjacent trials will increase over what will be observed if trials are distributed, provided the trials considered do not appear too early in the training series. Inhibition is assumed to develop rapidly and to reach a fairly stable level relatively early in training with massed trials. As a consequence, there will be a marked difference in level of inhibition taken from trial to trial in the earlier trials.

Whether the increased homogeneity of what is measured by a psychomotor test in later trials is associated with decreased correlation with other tests or criterion measures has not been adequately demonstrated. It has been found that the first 2 min. of testing on the Complex Coordination Test, i.e., the practice period preceding the regular 8-min. test, produces scores which consistently show greater correlation with criterion measures than the final 2- min. test period (2, p. 151). If the earliest test trials on psychomotor tests generally have the highest validity, then this supports the contention that the use of many different brief tests permits establishment of composites of high validity.

Purpose

The study reported here was designed to answer this specific question: Is the correlation between two psychomotor tests affected by the spacing of trials for the test performed first?

Procedure

Apparatus.—The first of the two psychomotor tasks used in this study was developed by the writer. The second task was adapted by the writer from the Rudder Control Test (2, p. 400 ff.; 4). The first test will be referred to in what follows as the Balance Test. The second test will be called the Six-Target Rudder Control.

The Balance Test consists of a board balanced at its center. The board is 48 in. × 7½ in. × 1 in. It is attached to a metal axle which turns within ball bearing mountings. The axle is supported by two uprights. It is 10 in. from the base of the apparatus. Attached to the sides of the axle are two heavy springs, one on the right and one on the left. These springs are attached to the base of the apparatus. They serve as snubbers. On the base of the apparatus, immediately below the ends of the 48-in. board, are two wooden blocks 5½ in. high. These stop downward movement of the board. On both sides of the board, 13½ in. from the center, are two wooden strips, 1½ in. × 7½ in. × ½ in., mounted at right angles to the long axis of the board. These are used as end stops for *S*'s feet. The board can move downward 3½ in. at these points. Essentially the apparatus consists of a board mounted on an axle so that it turns like a teeter-totter. Mounted on the base of the apparatus is a small box, 10 in. × 8½ in. × 6 in., with the 10-in. side parallel with the balance board. On top of this box are two rows of five lights each, parallel with the long side and with each other. One row has red lights, and the other, green lights. Attached to the axle is a set of sliding contacts which moves across a set of fixed contact points as the board travels back and forth. The circuits to the five green lights are made and broken in phase with the movement of the board. Perceptually, a green light moves to the right when the board is lowered on the right and to the left when the board is lowered on the left. The red light circuits are closed, in a random sequence, through a 40-point stepping relay. Whenever a given red light circuit is closed through the stepping relay, and the green light circuit with which it is paired is closed through the sliding contacts attached to the balance board, an electronic time delay mechanism is activated. If the mechanism is activated for .5 sec., the time delay mechanism discharges an impulse which advances the stepping relay. A new red light circuit is closed. Perceptually, the red light moves to a new position.

The *S* stands upon the balance board with the outer sides of his feet against the two wooden stops. The outer side of each foot is 13½ in. from the center of the board. The *S* is permitted to use his arms in any way he chooses except that he is not allowed to support his legs with his hands. He attempts to move the board and hold it balanced in such a fashion that the green

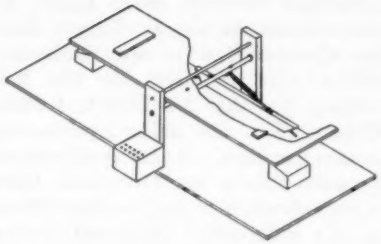

FIG. 1. Sketch of the Balance Test apparatus

and red lights are matched for .5 sec. The *S* attempts to match as many settings as possible in the given work period.

The task established by the Balance Test is one requiring counterbalancing movements of the legs and the trunk of the body. Even with the use of snubbing springs to reduce the difficulty of achieving momentary balance, the task is still a very difficult one. The task requires great effort on the part of *S*. The *S*s report fatigue, develop stiffness in the lower limbs, and sweat profusely while manipulating the balance board. A sketch of the apparatus is presented in Fig. 1. The sketch is not drawn to scale.

Basically the Rudder Control Test consists of a metal cockpit mounted in unstable equilibrium and a rudder which permits *S* to maneuver the cockpit. The task established by the apparatus requires counterbalancing movements of the lower limbs and shifting of the weight of the body. The Six-Target Rudder Control Test has a display panel of six green lights and a single white light. During the test period one of the green lights is illuminated at all times. The *S* attempts to move the nose of the cockpit in line with the green lamp which is lighted. When the nose is properly aligned with the green light, the white light comes on. When the nose of the cockpit and the green light are lined up for 5 sec., a time delay mechanism discharges and a stepping relay advances and the position of the lighted green lamp changes.

Training: Balance Test.—Two groups, each composed of 100 basic airmen trainees, were trained on the Balance Test. Both groups were given 80 settings to match. The first group (M) worked continuously. The second group (D) had a 1-min. rest following the first and the second five settings, and a 2-min. rest following each block of ten settings thereafter. During rest periods, *S*s stepped down from the apparatus and were seated.

Testing: Balance and Six-Target Rudder Control Tests.—Approximately 25 min. following the completion of the first 80 settings, both groups were presented 40 settings with no rest intervals. Approximately 15 min. later, both groups were given an 8-min. test period on the Six-Target Rudder Control Test. There was no rest interval during this test period. Two Group D *S*s, trained on the Balance Test, could not be tested on the Six-Target Rudder Control Test.

RESULTS

Performance curves for the Balance Test are presented in Fig. 2 and 3. During most of the training series, Group D performed at a level consistently superior to that attained by Group M. During the test series, when trials were presented without rest intervals, Group M was superior in performance to Group D. Differences between groups were not marked at any point in training or testing.

In the analysis of the results of the present study, the Balance Test training has been divided into two segments of 40 settings each. Also, there is a test period of 40 settings following training. These segments will be identified as 40_1, 40_2, and 40_3, according to their order of occurrence. Table 1 presents means, *SD*'s, and *t*'s for Groups M and D for these segments. Data for the 8-min. test period and for the first settings with the Six-Target Rudder Control Test are also given. The *t*'s for the differences between group means for the

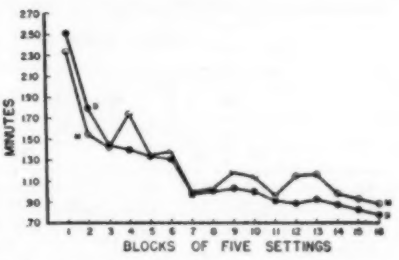

FIG. 2. Time, in minutes, required for matching successive blocks of five settings on the Balance Test for massed (M) and distributed (D) trained *S*s

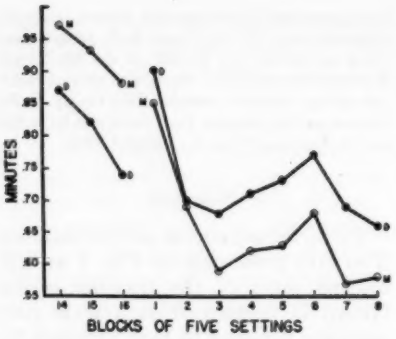

FIG. 3. Time, in minutes, required for matching successive blocks of five settings for groups previously trained with massed (M) or distributed (D) trials, and tested with massed trials. The last three points in Fig. 2 are shown for purposes of comparison.

40_1 segment and the total training series are not significant. The t for the difference between means based upon the 40_2 trials of the training series is significant at the 5% level of confidence, and is in favor of Group D. The t for the difference between group means for the test series is also significant at the 5% level of confidence. In this case, the difference is in favor of Group M. The difference between group means for total scores on the Six-Target Rudder Control Test is not significant. Mean time required for the first setting on the Six-Target Rudder Control does not differ significantly between groups. As far as differences between means are concerned, there is no indication of any residual effects of the differential treatment during training on the Balance Test.

Table 2 presents correlations of total and part scores for the training period and test period scores for the Balance Test with scores for the Six-Target Rudder Control Test. Included are t values for correlation differences between Groups D and M. As would be expected, all intercorrelation coefficients for the Balance Test are significant for both groups.

The conditions of presentation of training trials apparently affect the correlation of the Balance Test scores with the Rudder Control Test scores. The Group M scores do not correlate significantly with the scores on the Six-Target Rudder Control Test. The Group D scores for Trials 40_1 and 40_2 do correlate significantly with scores on the Six-Target Rudder Control

TABLE 1

MEANS, STANDARD DEVIATIONS, AND t's BETWEEN MEANS FOR THE TRAINING AND TEST PERIODS OF THE BALANCE TEST, AND FOR THE 6-TARGET RUDDER CONTROL TEST (6-TRC)†

	Training—Balance Test						Testing—Balance Test		Testing—6-TRC			
	40_1		40_2		Total		40_2		Time for Setting 1		8 min. Score	
	M	D	M	D	M	D	M	D	M	D	M	D
Mean	11.92	11.85	8.37	7.32	20.30	19.14	5.22	5.88	1.50	1.45	27.12	26.63
SD	5.20	5.11	3.28	3.68	7.27	8.09	1.94	2.53	1.22	1.36	13.44	17.54
N	100	100	100	100	100	100	100	100	100	98	100	98
tM_D–M_M	.10		2.13*		1.06		2.07**		.28		.22	

† M and D indicate massed and distributed practice groups. All means except 6-TRC 8 min. score means are stated in minutes. The 3 segments of training and testing on the Balance Test are designated 40_1, 40_2, and 40_3.
For $df = 200$: *$t = 1.97$ is significant at the 5% level of confidence.
**$t = 2.60$ is significant at the 1% level of confidence.

TABLE 2

INTERCORRELATIONS BETWEEN SCORES TAKEN AT TWO DIFFERENT STAGES OF TRAINING AND DURING A TEST SERIES ON THE BALANCE TEST AND WITH SCORES ON THE 6-TRC FOR D AND M GROUPS†

		Testing (40_3) Balance Test		Testing 6-TRC		Balance Test 40_3	
		D	M	D	M	D	M
Training (40_1 and 40_2) Balance Test	r	.72	.84	.34	.10	—	—
	t	2.13*		1.75			
Testing (40_3) Balance Test	r	—	—	.10	.04	—	—
	t			0.41			
40_1 Balance Test	r	.58	.48	.38	.17	.71	.44
	t	0.96		1.60		2.85**	
40_2 Balance Test	r	.70	.73	.27	−.04	—	—
	t	0.39		2.24*			

† Values of t are based on correlation differences between the D and M groups.
For $df = 100$: $r = .195$ is significant at the 5% level of confidence; $r = .254$ is significant at the 1% level of confidence.
For $df = 200$: *$t = 1.97$ is significant at the 5% level of confidence.
**$t = 2.60$ is significant at the 1% level of confidence.

Test. The 40_3 scores for Group D, however, do not correlate significantly with the Six-Target Rudder Control Test scores.

Massing of trials apparently affects the relationship between scores taken during the different segments of training. This can be noted in Table 3, which presents the values for t tests of significance for differences between r's for different stages of Balance Test practice as well as those for the Rudder Control Test scores. For example, the r for 40_1 and Rudder Control Test scores is compared with the r for 40_2 and Rudder Control Test scores. The difference between the r's for 40_3 scores with 40_1 and 40_2 scores for Group M is significant beyond the 1% level of confidence. The r for 40_2 scores and 40_3 scores is significantly greater than the r for 40_1 and 40_3 scores. The difference between the same pair of r's for Group D is not significant.

TABLE 3[2]

VALUES OF t FOR DIFFERENCES BETWEEN INTERCORRELATIONS WITHIN GROUPS OF GIVEN PART SCORES FOR BALANCE TEST AND 6-TRC†

X Variable	Y Variable: Testing—6-TRC		Y Variable: Balance Test—40_3	
	D	M	D	M
40_1 and 40_2	0.85	1.50	1.46	2.83**
40_1 and 40_3	2.07*	0.91		

† X variables from each row are paired with Y variables from each column for determination of each pair of compared r's.
For df = 200: *$t = 1.97$ is significant at the 5% level of confidence.
**$t = 2.60$ is significant at the 1% level of confidence.

[2] The values for t have been computed under the assumption of independent random sampling, which tends to reduce the size of the standard errors. Computation using a formula for related measures changes the significance only of the t for Group D for 40_1 and 40_2 with 40_3. The t becomes significant at the 5% level.

Continued training with changed conditions of presentation apparently reduces the correlation between the Balance Test and Six-Target Rudder Control Test scores. For Group D, the difference between the r for the Six-Target Rudder Control Test scores with 40_1 scores and the r for the Rudder Control Test scores and 40_3 scores is significant beyond the 5% level of confidence. The related pair of r's involving 40_1 and 40_2 and the Six-Target Rudder Control Test do not differ significantly.

Massing reduces the size of the r between early and late segment training scores. The difference between r's for 40_1 and 40_2 scores for Groups M and D is significant beyond the 1% level of confidence.

Discussion

The first question raised in the introduction to this report has to do with the effects of spacing of trials in the training of one task upon performance of a second, related task. Group M was trained with continuously presented trials throughout training. Group D had continuous practice only for the 40_3 segment. In terms of *time*, Group D had continuous practice 5.88 min. in a total of 25.02 min. Group M had 25.02 min. of continuous practice. If greater amount of practice spent under continuous conditions of presentation of trials is associated with a greater amount of residual inhibition, Group M should have a greater amount of such residual inhibition than Group D. On the Six-Target Rudder Control Test, however, there is no difference between group means for the first setting. The over-all mean for Group M is 27.12 settings and for Group D, 26.63 settings. The mean time for the first setting is 1.50 min. for Group M and 1.45 min. for Group D.

Although the over-all level of performance on the Rudder Control Test is not affected by the manipulation of spacing of trials during training on the Balance Test, the correlation between scores on the two tests is affected. This result and some of its implications will be considered. These implications have relevance for all the questions raised in the introductory section above. For this reason, this particular result will be discussed along with other changes in correlation observed.

Perhaps the changes in correlation relatable to conditions of practice can best be interpreted within the framework of the relatively simple and general theory suggested in the introductory section. It has been assumed that inhibition develops as a function of the amount of training and the relative degree of massing of training trials. The most relevant variables involved in the growth of such inhibition, or the nature of the changes in response associated with its growth, need not be considered here. It will be assumed further that there are individual differences in the rate of growth of inhibition. As inhibition develops, such individual differences become more prominent in determining between-*S* differences. The correlation between scores taken at different stages of practice is a function of the difference in the level of inhibition present, the r being larger when the level of inhibition is more nearly the same, decreasing as the difference increases. Distributions of practice data are consistent with the assumption that limits for growth of inhibition curves are approached relatively early in training trials whenever trials are massed. This would mean, in the present instance, that the greatest trial to trial differences in level of inhibition would occur relatively early in training for massed training *S*s. In the case of distributed practice, trial to trial differences in level of inhibition will be less marked. Such differences as exist, however, will persist from trial to trial for a greater number of trials.

The first two training segments, 40_1 and 40_2, involve distributed trials for one group and massed trials for the other. We should expect that the correlation between 40_1 and 40_2 will be lower for Group M than for Group D, since the difference in levels of inhibition present during 40_1 and 40_2 is assumed to be greater for Group M. The r for group M is .44; for Group D, .71. The difference is significant beyond the 1% level of confidence.

Since Group M had massed trials throughout training and testing and Group D had massed trials for 40_3 only, the Group M r between total score for the training series ($40_1 + 40_2$) and the test series (40_3) should be greater than the same r for Group D. The Group M r is .84 and the

Group D r is .72. The difference is significant beyond the 5% level of confidence.

In evaluating the correlations between the Balance Test and the Rudder Control Test scores, three facts must be kept in mind. Differences in performance of the two groups during the training and the test series with the Balance Test are never marked. The two groups do not differ significantly in performance during the whole training series. The performance of the two groups on the Six-Target Rudder Control Test is essentially identical. Finally, it must be remembered that the Six-Target Rudder Control Test was administered under continuous conditions of testing for 8 min. These testing conditions were more similar to those holding during the Balance Test phase for Group M than for Group D. These circumstances can all be conceived as operating against the appearance of the relationships observed.

There is nothing stated within the theory discussed above which enables any unequivocal prediction to be made with respect to the correlation of Balance Test scores with Six-Target Rudder Control Test scores. Since the correlation between parts of a test decreases as a function of the training time, or number of trials, separating the parts as well as the relative degree of massing of training trials, one hypothesis which is suggested is that correlation of scores in a given test with scores on any other test will decrease as a function of amount of training and relative degree of massing of trials. This hypothesis has been tentatively adopted as a part of the general theory employed in interpreting results of studies of the type reported here. According to this hypothesis, the r between Balance Test scores and Six-Target Rudder Control Test scores should be higher for Group D than for Group M. For both M and D groups the correlation of the Balance Test scores with the Six-Target Rudder Control Test scores should decrease with increased training on the Balance Test. As has already been noted, Group M r's for the Balance Test and Six-Target Rudder Control Test are not significant. Those for Group D are significant in three of the four instances considered. The nonsignificant r is observed when Group D shifts from distributed to massed presentations of trials.

With respect to the changes within groups, we should expect that the r's for Rudder Control scores and the indicated segment scores should arrange themselves in decreasing order of magnitude as follows: 40_1; $40_1 + 40_2$; 40_2; 40_3. The Group M r's arrange themselves as follows: .17; .10; −.04; .04. The Group D r's are ordered as follows: .38; .34; .27; .10. This is in agreement with expectations. The differences between the first and last r for Group D—.38 and .10—are significant beyond the 5% level of confidence. There is nothing in the results which seems to be inconsistent with the hypothesis examined above.

Any statement about the relationship of the present results to the general theory of testing must, necessarily, be very tentative. Since the Rudder Control Test has high validity as an aircrew selection test (2, p. 440 ff.), it has been used in the present study as a secondary validity criterion. When so used, the results obtained would seem to support the contention that early testing scores are more valid than later testing scores. The results are hence not inconsistent with the notion that a feasible method of constructing tests is to establish batteries consisting of many brief tests.

Summary

Two groups of 100 airmen basic trainees were trained on a Balance Test, one group with and one group without rest intervals. Following training, both groups were tested on the Balance Test without rest intervals. Following training and testing on the Balance Test, both groups were tested on a second test—a six-target form of the Rudder Control Test—without rest intervals.

Scores on the Balance Test for the group trained with distributed trials correlated significantly with the Rudder Control Test scores. Scores for the group trained without rest intervals did not correlate significantly with the Rudder Control Test. The correlations for the distributed training group decreased as training increased. The correlation of final part scores on the balance test with rudder control test scores is not significant for the distributed training group.

The correlation of initial scores with final scores on the Balance Test is lower for the continuously trained group than for the distributed training group. The correlation between part scores on the balance test for adjacent segments of training taken later in training is greater for the continuously trained group.

The results observed are interpreted within the framework of an inhibition theory. The relevance of the results to certain problems in psychomotor testing and motor theory are discussed.

(Manuscript received August 16, 1951)

References

1. Greene, E. B. An analysis of random and systematic changes with practice. *Psychometrika*, 1943, **8**, 37–52.
2. Melton, A. W. (Ed.) *Apparatus tests*. Report No. 4, Army Air Forces Aviation Psychology Program Research Reports. Washington: U. S. Govt. Print. Office, 1947.
3. Reynolds, B. Unpublished data.
4. Reynolds, B. A Six-Target Rudder Control Task. *Research Note* 50-3, Perceptual and Motor Skills Research Laboratory, Air Training Command Human Resources Research Center, Lackland Air Force Base, Texas, 15 June 1950.
5. Viteles, M. S. The influence of training on motor test performance. *J. exp. Psychol.*, 1933, **16**, 556–564.

A PROJECTIVE MEASURE OF NEED FOR AFFILIATION [1]

THOMAS E. SHIPLEY, JR.,[2] AND JOSEPH VEROFF [3]

Wesleyan University

McClelland *et al.* (5, 7) have developed what appears to be a valid method for measuring achievement motivation by means of a content analysis of the Thematic Apperception Test (9) or TAT stories. The TAT used here involves the writing of brief imaginative stories in response to certain pictures. Recently there have been attempts to extend the applicability of McClelland's content analysis to other needs. Roby (10) and Birney (3) have developed scoring systems for the need for security. The problem of finding a valid method of measuring the strength of the need for affiliation (*n Affiliation*) has arisen out of the attempts to study the need for security. Birney found that a group of college men whose need for security had been experimentally aroused by leading them to believe that they had scored poorly on a personality inventory showed significantly more imagery related to affiliation than a control group. The system that Birney used for scoring *n Affiliation*, however, did not seem general enough. The present study, therefore, was initiated to extend his work by arousing the affiliation motive more specifically and by finding a more generally applicable measure of it.

This paper will present the procedure adopted to arouse *n Affiliation* experimentally, the final scoring system for measuring *n Affiliation*, the results of applying the scoring system to experimental groups and to groups supposedly differing naturally in need for affiliation, and finally some additional indications of the validity of the measure of *n Affiliation* obtained.

Procedure

Two studies were designed to obtain a valid measure of *n Affiliation*. The first study was a comparison of the stories written in response to pictures by one group whose affiliation motivation had been experimentally aroused with stories written by a control group whose affiliation motivation was not specifically engaged. It was hoped that a valid measure could be developed by comparing the types of affiliation imagery produced by the two groups. The second study, designed to substantiate the validity of the measure obtained in the first study, was a comparison of the stories written by two groups whose affiliation motivation could reasonably be assumed to be of different intensity. In this study the affiliation motivation was assumed to be stronger in a group of college freshmen rejected from fraternities as compared to a similar group accepted into fraternities.

Experiment I.—The study concerned with the experimental arousal of *n Affiliation* was performed at the University of Pennsylvania. Two fraternities were selected which seemed most nearly the same in respect to social background, status on the campus, and size of actual membership. Both were invited to participate in research projects sponsored jointly by the United States Public Health Service and the Office of Naval Research. They were told that it was a study of group behavior and that $25 would be donated to the fraternity treasury for their cooperation.

Fraternity A was chosen arbitrarily as the experimental group (Aroused Group). From this fraternity 37 out of 60 members served as

[1] This report is the product of a series of exploratory studies into the nature of security motivation financed by grants from the United States Public Health Service under the general direction of Dr. David C. McClelland, to whom the authors are indebted. Data used in these studies have also been collected with the aid of grants from the Office of Naval Research for a study of achievement motivation which are also gratefully acknowledged.

[2] Now at the University of Pennsylvania.

[3] Now at the University of Michigan.

Ss. A sociometric test was given to arouse *n Affiliation*, and this was followed by the TAT from which it was hoped a measure of *n Affiliation* could be derived. Forty-five out of 60 members of Fraternity B (Control Group) were given a food preference test prior to the story-writing procedure.

Both groups were tested in their respective fraternity houses. Some members in both groups were excluded to adapt the *N*'s to the experimental procedures. The *Ss* were selected on the basis of their order of appearance in the room used for the testing procedure.

In the introduction to the sociometric procedure[4] *E* (TES) told the *Ss* that the following test was concerned with the problem of morale; that the Navy was interested in the friendship patterns of groups and how members of a group describe one another. The sociometric test for the Aroused Group was so arranged that each *S* was described by every other *S* from a list of 15 adjectives. These adjectives were: *aggressive, anti-social, argumentative, conceited, cooperative, entertaining, friendly, independent, intolerant, modest, self-assured, sincere, submissive, sympathetic*, and *timid*. On the top page of a booklet given to each *S* was a list of the adjectives arranged in alphabetical order. The *Ss* first rated these adjectives from 1 to 15, i.e., the rating 1 was to be given to that adjective which when ascribed to an individual would make him most attractive to the rater, etc. Then the first *S* stood up and the other *Ss* were instructed to write his number, 1, at the top of the second page, which also contained the same adjectives and to check the two adjectives on that page which best described him. There was one page listing the adjectives for each *S* present. This procedure was continued until all of the *Ss* present had stood up and been rated. Each *S* was also requested to rate himself when his turn for standing occurred. A separate list of names corresponding to the order of standing of each *S* was kept by all *Ss*, and from this list *Ss* were also requested to pick the numbers of those three individuals from the group whom they would choose as close personal friends. It was felt that these cues in the sociometric procedure—rating fraternity brothers, being rated by fraternity brothers, and choosing friends—would arouse *Ss*' *n Affiliation* to a greater intensity than would the cues ordinarily present in a fraternity group.

The TAT was introduced immediately following the sociometric test by a second *E* (JV). The *Ss* were told that both the USPHS and USN were interested in a research project concerned with creative imagination. The standard instructions (9) for the group form of the TAT were then presented. In brief, they were told that it was a test of creative imagination and that they were to write 5-min. stories in response to each of the five pictures to be projected on the screen before them. Each *S* was given a booklet containing five large sheets of paper. On each sheet appeared four questions. These questions were the standard ones that are customarily employed to insure complete coverage of all the elements in a plot or theme. The five pictures included four from the TAT (No. 6BM, 3BM 20, 10) and a fifth picture showing a group of young men seated around a table with another young man in the background. After they wrote their stories, *Ss* were told to write their numbers used in the sociometric test on the top of their story forms. At this point the group was told that both these tests were to be anonymous, and hence *Ss* were requested to destroy their pages listing the names corresponding to the numbers.

The Control Group wrote stories to the same five pictures after having taken the food preference test. They were told that contrary to their opinion, perhaps, the Navy was interested in the food likes and dislikes of young men. Each *S* was presented with two lists each containing 20 common foods. The *Ss* were told to check those foods on the first list which they would refuse to eat under most circumstances. On the second list they were to indicate by an L or a D whether they liked or disliked a particular food.

Experiment II.—This study was done at Wesleyan University. One month following the college's rushing program, 33 freshmen were randomly selected from all those who had been accepted into fraternities. These men constituted the Accepted Group. A Rejected Group was composed of 16 freshmen selected from those freshmen who had signified their desire to be pledged to a fraternity but had been rejected. In the judgment of the Dean of Freshmen and/or the college doctor each man who was finally selected for the Rejected Group had shown some indication that he had been disappointed by the rejection. Because of the limitation of information and the number of *Ss* available, no systematic attempt was made to match the *Ss* of these two groups on additional variables that might conceivably affect the affiliation motivation. However, a check indicates that the two groups were comparable with respect to minority-group membership and socioeconomic status.

These 49 freshmen were sent a letter from the Dean of Freshmen requesting them to meet in a lecture hall to take a test that had not been in-

[4] We are indebted for this procedure to Dr. Roger W. Heyns and Dr. John W. Atkinson of the University of Michigan.

cluded with the other tests given when they had matriculated. These *Ss* were given only the written TAT with no previous arousal as in Exp. I. The standard instructions were used with the same pictures shown in Exp. I, except that a sixth picture from the TAT (No. 14) was included at the end.

Scoring System for *n Affiliation*

The stories of the two groups in the arousal experiment were compared for types of imagery related to affiliation. The kind of imagery that seemed to discriminate the Aroused Group from the Control Group was brought together in a scoring system adopted to measure *n Affiliation*.

The subcategories employed for scoring the stories are similar to those successfully adopted by McClelland *et al.* for achievement (**7**) and hunger (**2**). That is, the subcategories seem to have general applicability for a wide variety of motives. The problem here was to define an affiliation-related story so that one could look for the presence of the subcategories which have been shown to reflect the motive strength for other needs.

Imagery.—Affiliation imagery is considered to be present where there is an objective statement in the story that a person is separated from another and is concerned about it or concerned about possible separation. This means that a story to be scored for *Imagery* should contain at least one of the following types of imagery: (*a*) concern with rejection, being jilted, "stood up," left out, outcast or ignored; (*b*) concern with loneliness, being without former friends or relatives, including mere mention of the word "lonely"; (*c*) concern with physical departure (e.g., negative affective concern over the death of a loved one); (*d*) concern with psychic separation (i.e., a quarrel, fight, or disagreement); (*e*) concern with no reciprocal love (i.e., one loves another, and is concerned because the other does not love him); (*f*) reparation (i.e., seeking forgiveness, repenting, or changing one's ways to preserve an interpersonal relationship).

A clue which frequently but not always indicates concern over separation is the introduction into the story of some person not in the picture used to elicit the story. This lead, which by itself yields too narrow a definition of affiliation imagery, has been followed up more extensively by Birney (**3**).

Unrelated Imagery.—The imagery in a story is considered to be unrelated if there is no mention of an affiliative loss or separation as defined above. Once a story has been scored *Unrelated Imagery* it is not, of course, scored for any of the subcategories below.

The following subcategories are scored if there is affiliation imagery present in a story. Each subcategory can be scored only once per story. Following each definition a brief example is given.

Need.—When there is a statement of a desire for the recovery, maintenance, or attainment of a friendly or loving relationship, *Need* is scored. Not scored *Need* are statements of a need for help, money, respect, etc., from a person. Some indications of a statement of *Need* are: "dreaming of," "prays for," and "desires his return." *Example*: Her son has been sent to Korea. She wants him to return quickly.

Instrumental Activity.—When there is a statement that someone in the story plans or acts to preserve or gain a friendly or loving relationship, the story is scored for *Instrumental Activity*. *Example*: He is trying to get the girl's mother to permit her to marry him.

Goal Anticipation.—This category is scored when someone is anticipating the recovery, loss, or gain of the friendly or loving relationship. *Goal Anticipation* is also scored when someone in the story anticipates the results of the recovery, loss, or gain. *Example*: He feels sure that his girl will consent to marry him.

Obstacle.—When the individual concerned about a relationship has personal shortcomings, such as jealousy or bad character, which stand in his way of preserving, obtaining, or recovering the personal relationship in question, the story is scored for *Obstacle*. The obstacle can also be in the environment. When something or someone (other than the desired person) stands in the way of the individual concerned, the story is also scored for *Obstacle*. *Example*: Her fiancé was killed in a train wreck.

Affective Goal State.—A story is scored for *Affective Goal State* when there is a statement of affect (positive or negative) over the attainment, preservation, or loss of the goal relationship, or when there is a statement of affect over any activity involved in gaining, preserving, or losing the goal relationship. Affect may be inferred from statements of marriage, making up, etc., culminating a relationship that had been scored for *Imagery*. *Example*: When he heard that Sally had jilted him, he wept and tore his hair.

Thema.—When the main plot of a story or one of two equally predominant plots in a story is concerned with the gain, preservation, or loss of a friendly or loving relationship, the story is scored for *Thema*. *Example*: A girl is slumped over her bed, weeping. She feels that there is little to live for since her father has just been reported killed in Korea. She wants him to return but realizes that this is now impossible. She will always miss him and resent the fact that he had to be called up for such service.

It should be pointed out that the scoring system is fairly easy and highly reliable. Reliability has been checked by two different methods. First, two scorers were compared for *Imagery* and subcategory agreement. For 294 stories there was 91% agreement over *Imagery* and an average of 88% agreement over the subcategories. The index of agreement was computed by dividing twice the number of agreements of a given category by the sum of the number of times the category was scored by the two scorers. Secondly, the reliability was checked by correlating over-all *n Affiliation* scores given to 49 *S*s by two scorers. The *n Affiliation* scores were calculated algebraically for each *S* by summing one point for each appearance of *Imagery* and any of the subcategories in his stories, and subtracting one point for each appearance of *Unrelated Imagery*. This inter-scorer product-moment correlation was .93, indicating a fairly high stability of an *S*'s score obtained by different scorers.

Results

The scoring system was applied to the stories of the Control and Aroused Groups. Two *E*s scored all the stories not knowing from which group a given story came. In the case of disagreements the final scoring was the result of a joint decision by the two scorers.

For each *S* in the arousal study the number of stories containing a particular category was determined. This number was converted into a percentage by dividing by five (the total number of stories written by each *S*). Therefore, an *S* could receive a percentage score of 0, 20, 40, 60, 80, or 100. These percentage scores were then converted into degrees by the *arc-sin* $\sqrt{percentage}$ transformation procedure. This was done in order to normalize the distribution of scores so that *t* tests might be applied to the data.

Table 1 contains the results of the category comparisons of the Control and Aroused Groups. Five out of the seven categories appear significantly more often in the stories of the Aroused Group than in the stories of the Control Group. Only *Unrelated Imagery* and *Goal Anticipation* appear more frequently in the Control Group, but only the difference for *Unrelated Imagery* is significant.

To check upon the validity of the scoring measure of *n Affiliation*, the *E*s scored the stories of the Accepted and Rejected Groups in the same manner as they scored the stories of the Aroused and Control Groups, and the same statistical treatment was employed. Table 2 reports the results of this study.

TABLE 1

Mean Number of Stories Containing Imagery and Subcategories Written under Aroused and Control Conditions

	Aroused Group $N = 185$ (37 *S*s×5 stories)	Control Group $N = 225$ (45 *S*s×5 stories)	Mean Difference	t^* $N = 82$	p^{**}
Imagery	3.35	2.53	.82	3.21	<.01
Unrelated Imagery	1.65	2.47	−.82	3.21	<.01
Need	2.14	1.07	1.07	3.56	<.001
Instrumental Activity	1.08	.62	.46	1.99 at	.05
Goal Anticipation	.41	.47	−.06	.37	>.50
Obstacle	2.16	1.56	.60	2.22	<.05
Affective Goal State	2.43	2.07	.36	1.62	<.11
Thema	2.65	2.15	.50	2.12	<.05

* Although the raw data are presented in the first three columns, the actual *t*-tests and resulting *p* values are based on an analysis of percentages that have been transformed by the arc-sin $\sqrt{percentage}$ procedure. The degrees of freedom for the *t*-tests, of course, are based on the number of subjects and not the total number of stories.

** These and all subsequent *p* values are for both tails of the distribution.

TABLE 2

MEAN NUMBER OF STORIES CONTAINING IMAGERY AND SUBCATEGORIES WRITTEN BY COLLEGE MEN ACCEPTED AND REJECTED BY FRATERNITIES

	Rejected Group $N = 96$ (16 *Ss* × 6 stories)	Accepted Group $N = 198$ (33 *Ss* × 6 stories)	Mean Difference	t* $N = 49$	p
Imagery	3.69	2.61	1.08	2.37	<.05
Unrelated Imagery	2.31	3.39	−1.08	2.37	<.05
Need	2.06	1.33	.73	1.63	<.11
Instrumental Activity	1.69	1.06	.63	1.73	<.10
Goal Anticipation	1.25	.67	.58	1.08	<.30
Obstacle	2.88	2.15	.73	1.83	<.10
Affective Goal State	2.94	1.91	1.03	2.40	<.05
Thema	3.38	1.97	1.41	3.62	<.001

* Again the reader is reminded that although the raw data are presented in the first three columns, the actual t tests and resulting p values are based on an analysis of percentages that have been transformed by the arc-sin $\sqrt{\text{percentage}}$ procedure. The degrees of freedom for the t tests, of course, are based on the number of subjects and not the total number of stories.

They are not so significant as those found in the arousal study. Nevertheless, three of the seven categories appear significantly more often in the stories of the Rejected Group at or beyond the .05 level of confidence. Three other categories in this study (*Need*, *Instrumental Activity*, and *Obstacle*) differ at approximately the .10 level but in the same direction previously obtained for the Aroused vs. Control Groups comparisons. Again *Goal Anticipation* did not differ significantly in either direction, while *Unrelated Imagery* appears significantly more often in the Accepted Group.

Perhaps it should be pointed out that because the various subcategories are not completely independent of the *Imagery* category, the p values obtained from the t tests are not accurate. However, the fact that the general procedure has been repeated at least once with confirmation of the direction of changes lends additional confidence to the findings.

The over-all effectiveness of the scoring system was tested by computing total *n Affiliation* scores for each *S* in all groups. For a given *S*, +1 was counted for each appearance of *Imagery* and for any appearance of the subcategories with the exception of *Goal Anticipation*, the one category which did not successfully discriminate between the two groups in either study. Also −1 was counted for any of the stories scored *Unrelated Imagery*. This category appeared significantly less often in the Aroused Group for the first study and in the Rejected Group for the second study. An *S*'s final *n Affiliation* score was the algebraic sum of these points. In Table 3 the significance of the differences between the mean *n Affiliation* scores

TABLE 3

MEAN *n Affiliation* SCORES FOR AROUSED VS. RELAXED GROUPS AND FOR REJECTED VS. ACCEPTED GROUPS

	Experimental Arousal Study		Fraternity Rejection Study	
	Aroused Group $N = 37$ (5 stories each)	Relaxed Group $N = 45$ (5 stories each)	Rejected Group $N = 16$ (6 stories each)	Accepted Group $N = 33$ (6 stories each)
Mean	12.11	7.51	14.31	7.64
σ_m	1.04	.99	2.00	1.37
Mean diff.	4.60		6.67	
$\sigma_{Mdiff.}$	1.47		2.46	
t	3.13		2.71	
p	<.01		<.01	

of the Aroused Group compared to the Control Group and of the Rejected Group compared to the Accepted Group are presented.

The mean *n Affiliation* score of the Aroused Group is significantly higher than the mean *n Affiliation* score of the Control Group, as is also the case for the Rejected Group as compared to the Accepted Group ($p < .01$ in both cases).

Discussion

Theoretical considerations.—McClelland has suggested a two-factor theory of motivation (6). He says that there may be two types of motives—one characterized by approach behavior and the other by avoidance behavior. He notes that there are behavioral indications that there are men whose achievement motivation is characterized by the hope of success and other men whose achievement motivation is characterized by the fear of failure. It has been shown (6) that the TAT stories of the men presumably motivated primarily by the hope of success contain achievement imagery with positive anticipation of the goal ($Ga+$), while men presumably motivated primarily by the fear of failure write stories containing achievement imagery with negative anticipation of the goal ($Ga-$). Other tentative indications of the two-factor theory applied to the achievement motive may be found in Atkinson (1) and McClelland and Liberman (8).

One may likewise argue for two aspects of affiliation motivation: (*a*) seeking affiliation because of the pleasant stimulus reward value of the affiliative relationship (approach behavior); and (*b*) seeking affiliation because of the painful stimulus value of rejection. With this theoretical orientation the present scoring system can be viewed as measuring predominantly that aspect of *n Affiliation* characterized by fear of rejection, i.e., the type of imagery scored is separation from objects of affiliation—primarily a statement of deprivation. Nevertheless, there are certain characteristics of the scoring system which may be related to approach behavior, such as positive affect or positive instrumental activity. However, these positive aspects of the affiliation imagery were not isolated in the present study because their incidence did not permit statistical treatment.

The results obtained may be readily understood by the deprivation theory of motivation, i.e., that behavior is motivated by the avoidance of a painful stimulus. Hence the Aroused Condition of this experiment probably threatened the men temporarily with possible deprivation of affiliation, and the rejected freshmen were in fact so deprived. Therefore, the separation imagery reflects the heightened deprivation state in the two studies. Although the results follow from this theoretical orientation, a two-factor theory of motivation has been discussed because the writers believe that a future experiment specifically designed to arouse the hypothesized approach aspect of the affiliation motive would present an opportunity to isolate a type of imagery different from the separation imagery herein presented. If such a factor should be isolated, different behavioral correlations would be expected as in the case of the achievement motive.

Validity of the n Affiliation measure.—The validity of the scoring system as a measure of *n Affiliation* primarily in its fear of rejection aspect can be indicated by two sets of data:

1. Rejection from a fraternity after one has indicated his desire to pledge must certainly be considered a rejection situation for freshmen in college. When rejected freshmen's stories were

compared to accepted freshmen's stories, the rejected freshmen should show more *n Affiliation* imagery than the accepted freshmen since all of these men had been turned down by at least one peer group fairly recently and had shown some concern about it. The results show that the Rejected Group as a whole did write more stories with fear of rejection imagery than the Accepted Group, although the category differences were not so great as in the arousal study. It was to be expected that the experimental study should show greater differences than those obtained in the fraternity rejection study. In fact, it may seem surprising that differences were obtained in the fraternity rejection study since one month had passed since the freshmen had been rejected by the fraternities before the testing was done. Also it should be noted that these boys in the fraternity rejection study probably perceived the testing situation as connected with Wesleyan's admission policy and not connected with the psychological experimentation done at the University. One might expect, therefore, that their affiliation motivation might be overshadowed by situationally aroused achievement motivation. Despite these two factors working against the possibility of discriminating the difference in the intensity of *n Affiliation* between the two groups, significant differences were found between the stories of the two groups. Consequently, not only has the validity of the scoring measure of *n Affiliation* been confirmed, but also a surprising sensitivity of the measure has been demonstrated.

2. It will be remembered that a sociometric procedure had been used to arouse *n Affiliation* in the Aroused Group. With the sociometric results the *E*s were able to obtain two different measures of popularity (popularity poll and a measure of social approval) to which the *n Affiliation* scores can be related. Inasmuch as it is maintained that the scoring system measures predominantly the fear of rejection aspect of *n Affiliation*, one would expect that an *S* who had a high *n Affiliation* score should be unpopular. An *S* who has been rejected by his group or considered unpopular should have a high *n Affiliation* score. If this hypothesis is borne out, there would be further confirmation of the validity of the *n Affiliation* measure. Moreover, the differences between the Accepted Group and the Rejected Group, which may be considered a popular and an unpopular group, respectively, would also be confirmed.

The first measure of popularity was the number of times an individual was chosen by his fraternity brothers as a close personal friend. The correlation between the number of votes for each *S* and his *n Affiliation* score computed by Kendall's *tau* rank-order technique[5] (4) was $-.27$ ($p < .05$). If one *S* who had the greatest number of votes and the highest *n Affiliation* score is not considered in the correlation, *tau* becomes $-.40$ ($p < .01$).

The second measure of popularity or social approval was more complicated. A popularity score was computed in the Aroused Group by summing the ratings given each of the two traits used to describe each *S* by every other *S*. The sociometric procedure demanded that each *S* rate 15 adjectives from 1 to 15 as described in the procedure. Furthermore, each *S* selected two adjectives to describe every other *S*. What this popularity or social approval score indicates, therefore, is the extent to which the adjectives used to describe a given *S* were favorable or unfavorable. The lowest absolute score obtained by an *S*, as computed by summing up each of the ratings given each adjective to describe that *S*, indicates that this *S* is the most popular or has the characteristics most approved. The *tau* obtained by correlating the *n Affiliation* score rankings with these popularity score rankings was $-.33$ ($p < .01$). Here again popularity is negatively correlated with *n Affiliation*. Both the first popularity correlation and the differences found between the Accepted and Rejected Groups in the fraternity rejection study are confirmed.

Nature of the arousal condition.—Since there is evidence that the scoring system measures predominantly the defensive aspect of the affiliation motive, or a deprivation state, one must consider the sociometric arousal situation in this light. What was there about this arousal situation that would elicit more imagery reflecting either an increase in the fear of rejection type of affiliation motivation or a heightened state of affiliation deprivation? The cues used to arouse the motive were describing others on personality traits, being described by others, and choosing personal friends. This situation was sufficiently ambiguous at the outset to raise doubts about an *S*'s final rating and to allow him to anticipate the possibilities of rejection. Although an *S* may have been overtly accepted by other *S*s in the group outside the experimental situation, in this covert situation, many *S*s would probably anticipate rejection. Also it will be recalled that *S*s were not informed specifically that resulting

[5] A rank-order correlation was used because of lack of confidence in the meaningfulness of a cardinal scale for the measures to be correlated. The *tau* rank-order technique was used in this and in the following computation because of the large number of ties both in the *n Affiliation* score and in the popularity vote frequencies.

description of *Ss* and the popularity of *Ss* would not be reported. Perhaps this is one reason why *Ss* in the Aroused Group wrote more stories with separation imagery than the Control Group.

Summary

Two studies were conducted to obtain a valid measure of the affiliation motive by scoring stories written in response to pictures. In one study 37 college fraternity brothers wrote stories in response to pictures after their affiliation motivation had been aroused by a sociometric procedure. A control group of 45 college fraternity brothers wrote stories after taking a food preference test. In the second study 16 college freshmen rejected from fraternities and 33 college freshmen accepted by fraternities wrote stories in response to pictures.

A scoring system for the affiliation motive was developed. The basic imagery of the stories scored by the system centered about the separation from affiliative objects. Applying the scoring system to the stories in the two studies yielded the following results:

1. In the first study, five out of the seven categories of the scoring system appeared significantly more times in the stories of the group aroused by the fear of rejection compared to the stories of the control group.

2. In the second study, three out of the seven categories of the scoring system appeared significantly more times in the stories of the rejected freshmen than they appeared in the stories of the accepted freshmen. Three other categories approached significance in the same direction.

3. An over-all measure of *n Affiliation*, derived by combining the categories that discriminated the two groups in at least one study, showed a significantly higher mean score for the group aroused in the first study and for the rejected freshmen in the second study.

An additional indication of the validity of the scoring measure of *n Affiliation* is that significant negative correlations exist between the over-all measure of *n Affiliation* for the *Ss* in the group aroused in the first study and two different measures of popularity obtained from the sociometric procedure.

Theoretical considerations about the nature of the affiliation motive are discussed.

(Manuscript received August 25, 1951)

References

1. Atkinson, J. W. A projective measure of achievement motivation in the interruption of tasks experiment. Unpublished Ph.D. thesis, Univer. of Michigan, 1950.
2. Atkinson, J. W., & McClelland, D. C. The projective expression of needs. II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.
3. Birney, R. C. The effect of threat on thematic apperception. Unpublished honors thesis, Wesleyan Univer., 1950.
4. Kendall, M. G. *Rank correlation methods.* London: Charles Griffin, 1948.
5. Lowell, E. L. A methodological study of projectively measured achievement motivation. Unpublished master's thesis, Wesleyan Univer., 1950.
6. McClelland, D. C. *Personality.* New York: William Sloane Associates, 1951.
7. McClelland, D. C., Clark, R. A., Roby, T. B., & Atkinson, J. W. The projective expression of needs. IV. The effect of the need for achievement on thematic apperception. *J. exp. Psychol.*, 1949, **39**, 242–255.
8. McClelland, D. C., & Liberman, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.
9. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
10. Roby, T. B. The effect of need for security on thematic apperception. Unpublished master's thesis, Wesleyan Univer., 1948.

DEGREE OF CONDITIONING OF THE GSR AS A FUNCTION OF THE PERIOD OF DELAY[1]

CARROLL T. WHITE

Human Factors Division, U. S. Navy Electronics Laboratory

AND

HAROLD SCHLOSBERG

Brown University

Recent theoretical discussions have placed considerable emphasis on the effect of various intervals of delay on the rate and extent of conditioning. The experimental data are in essential agreement as to the general form of the function; approximately a half second is the optimal interval between conditional and unconditional stimuli, and the gradient of effectiveness falls off on either side of this interval. But all of the previous experiments have been done on skeletal reflexes. Wolfle (**18, 19**) used hand withdrawal, Bernstein (**1**), Reynolds (**13**), and Kimble (**9**) used the eyelid reflex, and Kappauf and Schlosberg (**8**) used breathing changes in the white rat.[2] It should be noted that the latencies of the unconditional reflexes involved in these studies are all short, in the neighborhood of a tenth of a second. Hence it makes little difference in actual results whether the delay period is measured from S_c to S_u, or from S_c to R_u.

Hull's (**7**) theory assumes that the significant temporal relationship is between S_c and R_u; that is, the stimulus trace of the to-be-conditioned stimulus must be simultaneous with the unconditional response. On the other hand, Hilgard and Marquis were thinking of a stimulus-stimulus interval or more directly, an R_c–S_u relationship, when they suggested that the optimal delay interval was longer than the latency of the conditional reflex involved, thus permitting the CR to anticipate the S_u (**5**, p. 162). They had this generalization in mind when they stated that the most favorable interval for a long-latency response might differ from that for the short-latency responses. In fact, they pointed to the need for a study of such a long-latency response, mentioning the GSR as an example.[3] It is surprising that nobody has performed this obvious experiment, for it has considerable theoretical importance, and offers no great technical difficulties. The present paper reports an experiment based on their suggestion.

METHOD

The GSR was conditioned to light, with an electric shock as the S_u. A synchronous motor-driven timer system made it possible to work with any desired interval between the onset of the light and the shock. This timer was set up so that both stimuli would terminate simultaneously after having been presented together for a period of 1.5 sec. The intensity of the shock was held constant at 150 v., .1 ma., throughout.

[1] This research is reported in greater detail in White's thesis (Sc.M.), under the same title, which is available in the Brown University Library.

[2] We have not distinguished between delayed and trace conditioning, for the onset of the S_c is probably the important thing, especially at short intervals of delay.

[3] Hilgard (**6**) has more recently re-emphasized this same point.

The *S* was seated in a small, sound-resisting closet adjoining the experimental room. This closet was illuminated by a shaded 25-w. bulb which was not visible to *S*. An electric fan served the double purpose of circulating the air and providing a source of noise within the closet to help mask any outside distractions. The S_c was produced by a small 12-v. (DC) bulb behind a circular piece of milk-glass 2.5 in. in diameter which was located 4 ft. in front of *S* and 1 ft. above eye-level. The measured brightness of the light stimulus was 70.5 millilamberts. The S_u was delivered through electrodes which were placed on the right wrist, and the GSR "pick-up" electrodes were attached to the palmar surface of the left hand.

The response end of the apparatus consisted of an AC resistance-capacitance bridge circuit, a system for amplifying and rectifying the output of the bridge, and a recording system. The process of balancing the bridge circuit was greatly facilitated by the use of a small cathode-ray oscilloscope which was connected to the bridge output. The recording itself was made by an ink-writing system activated by the movement of a permanent-magnet type loudspeaker element.

Seven groups of ten *S*s were utilized in this experiment. Six of the groups were conditioned with S_c–S_u intervals of 0, ¼, ½, 1, 2, and 4 sec., respectively, and the remaining group was used as a control to determine the degree of pseudo-conditioning obtained in the situation. Each *S* in the six experimental groups was presented with five pairings of the light and the shock, their onsets separated by the appropriate interval. There was a minimum of 1 min. between each of these training trials. Following these five paired presentations *S* was presented with the light alone for five trials, also with at least 1 min. between presentations. Finally, *S* was presented with the shock alone, the intensity and duration of which were the same as in the training series. The procedure for the control group differed from the above only in that the light and shock were never paired, *S*s being presented with the same number of stimuli in the same total time as in the experimental groups, but with these stimuli equally spaced (approximately 30 sec. apart) and the light and shock presented in random order. The responses to the light-alone series (extinction) and to the shock-alone were recorded for all groups.

Results

The quantification of the GSR is a knotty problem at best; in this experiment it was further complicated by the use of alternating current, which involves changes in both resistance and capacity. Hence it seemed best to use relative, rather than absolute values. The primary data consisted of measured deflections on each of five extinction trials, and on a final shock. Each CR during extinction was converted to a percentage of *S*'s response to the final shock, and all computations were carried out with these percentages.

Table 1 presents the means for each interval group on each extinction trial. By averaging the rows it is possible to get an over-all measure of strength of conditioning for each period of delay. These values are plotted in Fig. 1. Since the distributions are skewed, the median percentages are also shown. Both mean and median show the same general picture; a marked optimum with a delay period of ½ sec., and a sharp drop on either side of the optimal interval. There is a suggestion that the ¼- and 1-sec. groups were better than the sensitization or pseudo-conditioning control group, but the difference is not significant. The ½-sec. group is significantly superior to all other groups at well beyond the .01 level of confidence when Wilcoxon's "Method of Non-

TABLE 1

Mean Values of Ten Responses on Extinction Trials after Conditioning at Various S_c–S_u Intervals

S_c–S_u Interval (Sec.)	Extinction Trials					Grand Mean
	1	2	3	4	5	
0	58.7	44.2	38.6	31.4	19.4	38.5
¼	77.9	50.1	48.9	47.2	39.7	52.8
½	99.3	83.0	65.9	66.4	58.7	74.7
1	49.0	48.7	53.9	53.0	44.4	49.8
2	39.6	36.1	27.5	25.2	31.1	31.9
4	46.3	41.4	39.8	50.6	40.5	43.7
Control	47.9	42.2	40.4	46.3	35.0	42.4

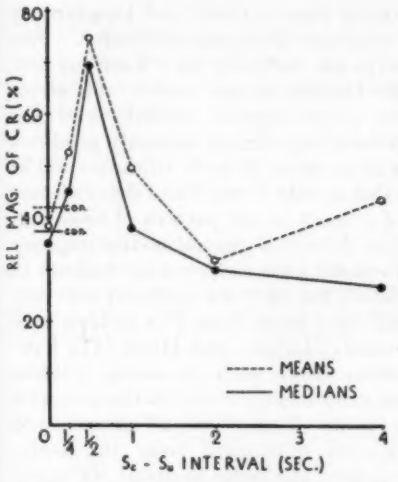

FIG. 1. Conditioning as a function of the length of the interval of delay. The horizontal lines (con.) on the ordinate represent results from a psuedo-conditioning control group.

Paired Replicates" (17) is used. Since there might be some objection to basing conclusions on percentages, a still more conservative test was made of this difference. The original data were examined to see how many CR's in each delay-group actually exceeded the responses to *S*'s terminal shock. Fourteen out of the 50 responses in the ½-sec. group satisfied this criterion, but only 2–4 did so in the other groups. Chi-square shows this difference to be far beyond the .01 level of confidence.

The data were also examined for the number of small responses in each interval group. For purposes of analysis, a small response was defined as one that was less than half as large as the response to the final shock. As might be expected, relatively few (38%) of the responses in the ½-sec. group were small. The poorest group by this criterion was that with the 2-sec. delay, which gave 80% small responses. All the other groups, including the pseudo-conditioning control group, had 54–64% of small responses. The excess of small responses in the 2-sec. group, as compared to the control group, was significant beyond the 2% level of confidence (Chi-square).

Since the data plotted in Fig. 1 are based on amplitudes of response during five extinction trials, they represent a hybrid measure of strength of conditioning. A more straightforward measure would be the mean of each group on the first extinction trial. When this is plotted (not shown), the curve is very similar to that of the means in Fig. 1; the only difference is that the means for simultaneous, ¼-sec. and ½-sec. groups are each raised by about 20%. These three groups fall steadily during extinction, but the ½-sec. group still maintains its superiority on the fifth extinction trial. An examination of Table 1 will show that there is very little loss during extinction in the control, 1-, 2-, and 4-sec. groups. These consistent trends, although they won't bear statistical analysis, support the general conclusion derived from the plot of medians in Fig. 1; nothing but pseudo-conditioning is obtained with the GSR if the interval of delay is 1 sec. or more.

DISCUSSION

The curve representing amplitude of conditional GSR as a function of delay interval (Fig. 1) is very similar to those obtained for the conditional eyelid reaction (**1, 9, 13**) and hand withdrawal (**18, 19**). Detailed comparison among the curves obtained in these experiments is not justified; in fact, one should not take the form of any of the curves very seriously, since there is no evidence that the response measures have been made on an equal-unit scale. But it is clear that the optimal delay period of the

conditional GSR is near ½ sec., just as it is for hand-withdrawal and eyelid reaction. Yet the latency of the unconditional GSR, either to light or to shock, is of the order of 2 sec., while that for eyelid reaction is in the neighborhood of a tenth of a second.

This finding suggests that the basic processes involved in conditioning the hand withdrawal, eyelid reaction, and GSR are similar. It certainly throws considerable doubt on Mowrer's theory (12) that skeletal and autonomic conditioning are fundamentally different, since the first two reflexes are skeletal, and the GSR is one of the better examples of an autonomic function. The results are almost as hard on a drive-reduction theory of reinforcement (7). The GSR is sometimes considered a measure of tension. It has a latency of the order of 2 sec., and reaches its maximum about 3 sec. after the shock. This means that the best delay period, ½ sec., was one in which S_c occurred just before a marked *increase* in tension, which is scarcely in agreement with the drive-reduction theory.

Let us now turn to what Spence (15) has called the contiguity theories. The results of the present experiment offer no particular difficulties for the S–S version of this theory; whether we think of the association being formed between sign and significate, S_c and S_u, or even between neural cell assemblies (4), the theory would predict that the optimal S_c–S_u interval would be determined by time relations among stimulus-released neural processes, rather than responses. For these theorists the long latency of the GSR would be of no importance. But for Guthrie's (3) type of theory, which stresses contiguity of stimulus and response, the failure to find any differences in optimal delay period between short-latency and long-latency responses offers some difficulty. Perhaps the difficulty isn't a serious one, for Guthrie places considerable stress on proprioceptive stimuli, and the present experiment certainly produces a large mass of such stimulation. The GSR is only a small and delayed part of a much larger pattern of behavior. The details of general bodily response to shock have not been worked out in detail, but they are probably not very different from those to a sudden loud sound. Landis and Hunt (11) have shown that such a startle pattern actually begins with a lid closure, with a mean latency of 40 msec., and spreads downward over the body, reaching the knees in about 300 msec. There is some evidence for a second phase of activity, starting about 1 sec. after stimulation, and involving the diencephalon (2); the GSR is apparently part of this phase (10). With this mass of activity going on in the muscles, it is hard to see how Guthrie could predict that any particular delay interval would be superior to any other.

From this brief discussion it is clear that an S–S contiguity theory is the only one among the major learning theories that might predict the superiority of the ½-sec. delay period for conditioning the GSR. Perhaps this is to be expected. The experimental situation was one in which a stimulus sequence was emphasized; *E* maintained a constant temporal relationship between onset of light and shock, and recorded only one element of a complex response. Furthermore, no particular response had important consequences to *S*. The eyelid reflex, and to a lesser extent, Wolfle's hand withdrawal, share these characteristics. Hence all three situations favor the development of an adaptation to a stimulus sequence, which is

the essence of an S–S theory. Experiments in which responses, rather than stimuli, are under experimental control may well produce results in conformity with the predictions of an S–R theory. In short, there is still no reason to attempt to force all learning experiments into one theory (**14, 16**).

So much for the chief finding, which was the marked superiority of the ½-sec. interval. There was one minor result that deserves a comment. The 2-sec. group showed a marked suggestion of inferiority to all other groups, including the pseudo-conditioning control. The implications of this aren't clear. It is possible that 2 sec. is just long enough to permit *S* to "set" himself for the oncoming shock, thus minimizing its effect. There may also be some significance in the fact that the latency of the original GSR to light is 2 sec. Thus, the *S*s in this group were receiving a shock just as the original response to light was starting. Such a shock may have had a punishing or disrupting effect on this original response, depressing it below the level it would attain through reflex sensitization or pseudo-conditioning. Further study of intervals around 2 sec. might turn up something interesting.

Summary and Conclusions

Six groups of ten *S*s each were conditioned in a light-shock situation with various S_c–S_u intervals. In addition to these, a pseudo-conditioning group of ten *S*s was used. All *S*s in the experimental groups were given five presentations of light and shock in the appropriate temporal relationship, followed by an extinction series consisting of five light-alone presentations. This in turn was followed by a presentation of shock alone. The responses to the light during the extinction series and to the shock alone were recorded, and from these records measures of relative magnitude were computed for all of the responses to light.

From the results so obtained it was found that, of all the intervals used, the ½-sec. interval resulted in the greatest degree of conditioning. Marked decreases in the degree of conditioning were observed as the interval was varied from this point in either direction. Thus, the optimal S_c–S_u interval for conditioning an autonomic response is similar to that usually reported for conditioning skeletal responses. It was also found that there was a tendency for the magnitude of the responses to be depressed when the 2-sec. interval was used. The significance of these results was discussed; they seem to fit best into an S–S contiguity theory.

(Manuscript received August 21, 1951)

References

1. Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.
2. Davis, R. C. Motor effects of strong auditory stimuli. *J. exp. Psychol.*, 1948, **38**, 257–275.
3. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1937.
4. Hebb, D. O. *Organization of behavior: A neuro-psychological theory.* New York: Wiley, 1949.
5. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.
6. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
7. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
8. Kappauf, W. E., & Schlosberg, H. Conditioned responses in the white rat. III. Conditioning as a function of the length of the period of delay. *J. genet. Psychol.*, 1937, **50**, 27–45.
9. Kimble, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1–15.
10. Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
11. Landis, C., & Hunt, W. A. *The startle pattern.* New York: Farrar and Rinehart, 1939.
12. Mowrer, O. H. On the dual nature of learning: A re-interpretation of "condi-

tioning" and "problem-solving." *Harv. educ. Rev.*, 1947, **17**, 102–148.

13. Reynolds, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35**, 15–30.
14. Schlosberg, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44**, 379–394.
15. Spence, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
16. Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, **56**, 144–155.
17. Wilcoxon, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.
18. Wolfle, H. M. Time factors in conditioning finger-withdrawal. *J. gen. Psychol.*, 1930, **4**, 372–378.
19. Wolfle, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.*, 1932, **7**, 80–103.

MOTIVATIONAL PROPERTIES OF FRUSTRATION: I. EFFECT ON A RUNNING RESPONSE OF THE ADDITION OF FRUSTRATION TO THE MOTIVATIONAL COMPLEX[1]

ABRAM AMSEL AND JACQUELINE ROUSSEL

Newcomb College, Tulane University

This paper reports an attempt to test two hypotheses derived from assumptions relative to frustration. The assumptions are: (*a*) that frustration is a motivational condition, and (*b*) that strength of frustration varies with time in the frustrating situation. In this experiment, frustration is defined as a state which results from the nonreinforcement of an instrumental response which previously was *consistently* reinforced. Saying that frustration can be produced in this way does not, of course, exclude other possibilities, such as blocking of a strongly learned instrumental sequence by a physical barrier.

In a recently reported study, Rohrer (7) employed the term *frustration drive* to account for the reduction in strength of a bar-pressing response. He showed that when two groups of rats are conditioned equally strongly to bar-pressing, and each is subsequently extinguished twice under a particular degree of massing of trials, the more highly massed group takes fewer trials to the second extinction. It is our opinion that such findings do not require the introduction of the notion of a motivational frustration state to explain them. They can be handled by existing inhibition theory (3). Rohrer's results are not in our opinion a test of the motivational characteristics of frustration, nor were they probably intended to be. In order to show that frustration is a motivating condition it would seem necessary to demonstrate that its presence increases some aspect of behavior. It is not sufficient simply to show a decrease in strength of the response leading to frustration.

If frustration is to be conceived of as a need state, it would, then, seem reasonable to demonstrate, at the outset, that this state has the accepted properties of need states. Thus, Hull's treatment of motivation (3) would demand that adding frustration to a motivational complex increases the strength of some ongoing response through increase in generalized drive strength, provided that the frustration drive stimulus has not become connected to a response which is antagonistic to the criterion response (1).

This experiment was designed to test the following specific hypotheses stemming from the earlier assumptions: (*a*) when a running response has been maximally elicited under hunger motivation, the addition of frustration to the motivational complex will result in the establishment of a new and higher maximum running speed; (*b*) the facilitating effect of frustration on running speed will vary with variation in the amount of time spent in the frustrating situation just preceding that running.

METHOD

Subjects and Apparatus

Eighteen male albino rats of Wistar strain were *S*s in this experiment. Their ages were

[1] Supported by a grant from the Tulane University Council on Research.

between 100–120 days at the beginning of preliminary training.

The apparatus was a modified simple straight-away consisting of a starting compartment, two runways and two goal boxes. They were arranged in a straight line in the following order: starting box, Runway A, Goal box 1, Runway B, Goal box 2. The entire apparatus, except for Goal box 2, was 2.5 in. wide and 4 in. high (inside dimensions). The starting box was 9 in. long; Runway A was 3.5 ft. long; Goal box 1, 1 ft. long; and Runway B, 10 ft. long. Goal box 2 was $11 \times 7 \times 10$ in. Guillotine doors separated the starting compartment from Runway A, and Goal box 1 from Runway B. At the entrance to Goal boxes 1 and 2 were swing doors designed to be manipulable by *S*. The entire apparatus, with the exception of Goal box 2, was painted flat black. Goal box 2 was unpainted pine. The apparatus was covered with ¼-in. hardware cloth mesh. A small metal food cup was flush with the floor of Goal box 1. The food cup in Goal box 2 was a clear glass coaster.

The measurements of performance were made in Goal box 1 and Runway B. A running time measure was taken by the use of a Standard Electric Timer (.001 min.) and a microswitch arrangement. The switches were placed under hinged segments of the floor of Runway B so that the weight of the animal would trip them. The first switch was placed 1 ft. beyond Goal box 1, the second switch, 1 ft. from Goal box 2. The running time was, then, the time to traverse the middle 8-ft. portion of Runway B. The latency measure was taken with a 1/10-sec. stopwatch. This was, on each of the test trials, the elapsed time between the raising of the door of Goal box 1, and the activation of the first microswitch in the running-time circuit.

A separate swing-door and goal-box arrangement was employed during preliminary training to acquaint each *S* with the door-pushing response. This was of similar construction to the experimental apparatus, but was 4 in. wide and 12 in. long, and was painted gray.

Experimental Procedure

The entire experimental sequence required 52 days. This period was divided into three sections.

Preliminary training.—This was a period of 12 days designed to adapt *S*s to handling by *E* and to the apparatus. All *S*s, beginning at this stage, were maintained on a 22-hr. hunger schedule. Their daily food ration was 9 gm. of Purina dog chow checkers. On each of the first seven days the *S*s were handled by *E*. On Day 8, *S*s were put into the apparatus in pairs with all doors raised, for 10 min. per pair. On Day 9 each *S* explored the apparatus individually for 7 min. Days 10, 11, and 12 were spent in training each *S* to manipulate a swing door such as those at the entrance to the goal boxes. Each *S* was given six trials on each of these three days. A pellet of food was given after each entrance into the goal box. After these trials none of the 18 *S*s had any difficulty manipulating the door.

Measurement of maximum running speed under 22-hr. food deprivation.—For the next 28 days, Days 13 to 40, each *S* was run three times per day in the apparatus under the strong hunger drive. During these trials food was in both goal boxes. The details of procedure on an individual trial are as follows: (*a*) *S* put into starting box; (*b*) after 3 sec., door raised; (*c*) *S* traverses Runway A, enters Goal box 1, eats one puppy pellet of chow; (*d*) after 30 sec. in Goal box 1, door is raised; (*e*) *S* traverses Runway B into Goal box 2, time to run this alley is measured; (*f*) *S* consumes food pellet in Goal box 2, and is returned to individual carrying cage.

The *S*s were run in rotation, nine at a time, so that each of the three runs on a single day for any *S* was separated from the others by about 20 min.

After these 84 trials on 28 days the running speed of the *S*s seemed to have reached a stable maximum.

Measurement of latency and running speed under 22-hr. hunger with and without frustration.—During this 12-day period each *S* was run three times a day, a total of 36 trials. Of these, 18 trials were *frustration* trials and 18 were *reward* trials. The reward trials were identical with those in the previous period. The frustration trials differed from the reward trials in that *S* found Goal box 1 empty (no food or food cup), after traversing Runway A. Each *S*, therefore, served as its own control to test the effect of frustration on the level of the performance which immediately followed the frustrating event.

To test the second hypothesis, *S*s were split into three subgroups, matched on the basis of mean running time in the preliminary running trials. These will be termed Group 5, Group 10, and Group 30. They are differentiated by the amount of time spent in Goal box 1, during these 12 test days, *on frustration trials*. Group 5 remained in the frustrating situation for 5 sec.; Groups 10 and 30 were in the frustrating situation for 10 and 30 sec., respectively. All *S*s continued to remain for 30 sec. in Goal box 1 on reward trials.

During this test period, two performance measures were employed on every trial. The

running time measure (time to traverse the 8-ft. segment of Runway B) was taken as before. In addition, a latency measure was obtained. This was the time interval separating the raising of the door of Goal box 1 (after 5, 10, or 30 sec.) and the activation of the first microswitch. The *E* raised the door of Goal box 1 only when *S* was oriented in that direction. This meant that on some trials, the 5-, 10-, or 30-sec. interval in Goal box 1 was slightly exceeded.

In running equal numbers of frustration and reward trials in this test period the following arrangement of daily trials was employed to control for any possible effect of order of presentation: FRF RFR FFR RRF FRR RFF. These symbols represent the order of frustration and reward trials for six days (18 trials). On the next six test days, this sequence was repeated.

Results and Discussion

Test of two hypotheses.—Both means and medians were calculated. In all cases the general nature of the results is the same for both, except that means are higher and more variable than medians.

To evaluate the results of the test period, mean and median running time and latency scores were computed for each *S* under reward (R) and frustration (F) conditions. Each of these scores is an average of 18 test-trial measurements. Comparison of the R and F scores showed that for each of the 18 *S*s, median running time and latency measures are lower on frustration trials than on reward trials. The means show similar results with the frustration trials producing faster subsequent running than the reward trials for 16 of 18 *S*s. Seventeen of 18 *S*s had shorter mean latencies on F trials than on R trials. The median of individual *S* median reward running times was .048 min.; the corresponding frustration median was .041. The over-all means were .054 and .044 min. The over-all median R and F latencies were 1.65 and .65 sec.; the mean latencies were 4.86 and 1.69 sec. Since the raw data cannot be presented here, it should be indicated that for any given *S* the distribution of 18 frustration measurements overlapped very little, when at all, with the 18 reward scores. The first hypothesis—an increase in strength of performance with the addition of frustration to a motivational complex—is strongly supported.

In order to test the second hypothesis, the test performance of the *S*s must be considered according to whether they were in the 5-, 10-, or 30-sec. group. Table 1 shows running time and latency medians and means for each of the three groups under reward and frustration conditions. A simple analysis of variance was performed with the individual R–F scores to test whether the three

TABLE 1

Running Time and Latency Averages on Reward and Frustration Test Trials

Group	Running Time (Minutes)				Latency (Seconds)			
	Median		Mean		Median		Mean	
	R	F	R	F	R	F	R	F
5	.053	.042	.056	.045	1.35	0.65	4.13	1.87
10	.048	.036	.052	.041	1.30	0.45	4.25	1.33
30	.046	.041	.053	.046	3.75*	1.25	6.19*	1.90

* Since running times, but not latencies, were measured during the preliminary period, the groups could be equated only for running times. For this reason, it is not surprising that Group 30 deviates so greatly from the other groups in Median and Mean latency on reward test trials.

conditions of time in the frustrating situation resulted in different frustration effects on running time and latency. The *F* ratio for running time was 3.06; the latency *F* ratio was 1.12. With 2 and 15 *df*, the 5% level of confidence requires an *F* ratio of 3.68. We cannot, therefore, say that time in the frustrating situation was a relevant variable in this experiment, although the average vigor of response of the Group 10 *S*s on frustration trials is consistently greater than that of the other groups.

Possibly, the range from 5 to 30 sec. *in a frustrating situation* is one which produces relatively little change in the frustration effect. The performance of the *S*s under 22-hr. food privation may already have been so strong as to obscure differences in frustration which may have been present in the three groups. Certainly, our groups were small for this test. It would seem reasonable, on the basis of common sense and on the basis of the typical characteristics of labile states, to expect a very rapid building up of the frustration state (perhaps almost complete by 5 sec.) followed by a gradual falling off of its intensity. This falling off would have a faster rate outside of the frustrating situation than in it.[3] Since, in this experiment, there was no significant difference between 5, 10, and 30 sec. in the frustrating situation, we offer the speculation that the period from 5 to 30 sec. is a period of relative stability of frustration before the dissipation of the state begins.

Comparison of reward and frustration test trials with preliminary reward trials.—Figure 1 shows the median running time data for reward conditions on the initial pretest days, and for the R and F conditions during the test trials. It should be noted that the two test-trial curves represent the median running times of the same *S*s under both reward and frustration conditions, so that the first F trial was the first test trial, whereas the first R trial was the second test trial, and, so on in random arrangement of F and R. For this reason, at any point on the abscissa of the test-trial graph the probability of the F curve being lower than the R curve would be .5, if there were no frustration effect. Actually, the F curve is lower than the R curve at 17 of the 18 points, and at most of these points it is very much lower. Taking the appropriate term in a binomial expansion, we find that the probability of such a chance occurrence is less than .0001. This is further confirmation of our first hypothesis.

The F curve approaches an asymptote which appears to be about .010 min. lower than that of the initial R curve, and that of the test-trial R curve. This is quite striking when one considers the great stability of the initial running time curve before the introduction of the test trials. The test reward curve remains at about the same level as the pretest curve. The greater apparent variability in the test trial data than in the initial data can be attributed to the fact that the test curves were plotted from medians of individual trials while the initial curve was plotted from medians of each day's performance, and there were three trials on each of these days for each *S*.

Since there is, on the first four test trials, very little difference between the R and F curves, and since the difference which becomes evident by Trial 5 is due mainly to a decrease in running time on F trials, there might be some basis for arguing that the frustration effect is learned, and not simply the result of a sudden increase

[3] A pertinent experiment of Miller and Stevenson (5) reports an increase in agitated behavior of rats immediately following nonrewarded trials. This increased agitation was observed in the alley leading to the goal box where *S* had 10 sec. previously been frustrated (nonreward following consistent reward). In the Miller and Stevenson study, this frustration-motivated response was found to decrease in vigor when intervals longer than 10 sec. separated a frustration occurrence from the next run of *S*. By 2 min., the agitated behavior was no longer apparent. Here, time between nonreward and observation of agitated behavior was spent by *S* outside of the frustrating situation.

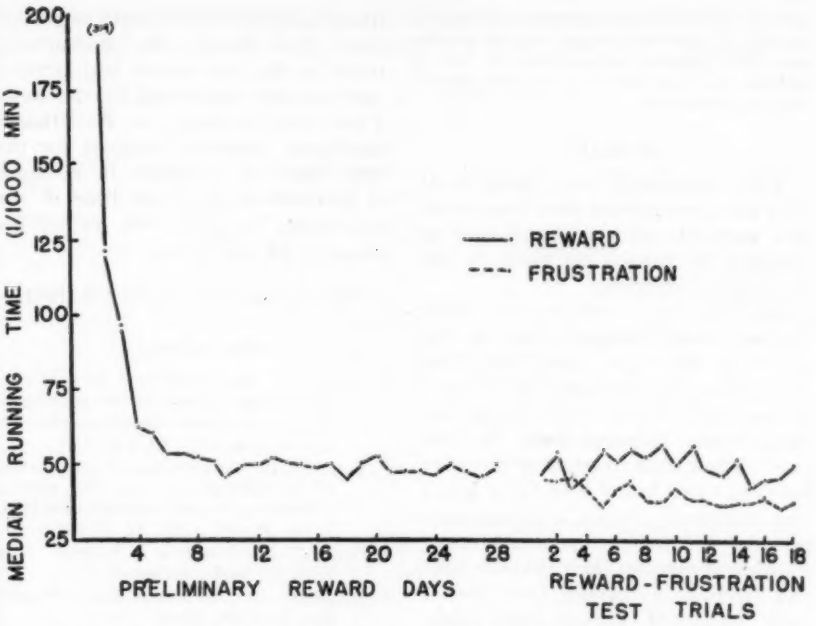

FIG. 1. Comparison of test-trial performance under reward and frustration conditions

in motivational strength. There is however an explanation of this part of the data which would not involve adopting this position. Early in the test period, frustration due to non-giving of food in the first goal box is minimized because of the strong secondary-reinforcing properties of that region built up during the initial trials; later, when the secondary-reinforcing power of this goal box extinguishes, the frustration effect is increased.

Another interpretation.—There is an interpretation of the present results which does not involve the introduction of frustration as an explanatory concept, i.e., it does not apparently necessitate attributing special properties to nonreinforcement.[3] This would hold that the running times in Runway B during the initial trials were actually depressed because of hunger reduction in Goal box 1 preceding those trials; and that from the beginning, the running time in Runway B would have been faster, and the final limit of performance would have been lower, if there were no eating just preceding the performance in the second runway. According to this interpretation, the increased vigor of performance which follows failure to reward a previously-rewarded response is due not to the presence of a new additional motivational component (frustration), but to the absence of a reduction in hunger tension which had consistently depressed performance before.

Actually, failure to reward in Goal box 1 changes the stimulus to running in Runway B on frustration trials, and should also, therefore, tend to *reduce* running speed on those trials. Also, there is evidence that introducing food immediately before a hunger-motivated response may actually increase the level of performance, presumably by increasing incentive motivation.[4] Both of these factors, if operating, would have an opposite effect to that suggested.

It is possible that all the factors mentioned are operating in this experimental situation, and that frustration-produced motivation is also operating. If special emergent properties are

[3] Suggested by Dr. I. E. Farber in a personal communication.

[4] Some support for this statement can be derived from certain of the early speed of locomotion experiments (**2, 6**). The whole question of the effect on vigor of performance of giving a small amount of reward immediately before a trial has recently been treated by Maltzman (**4**).

to be attributed to nonreinforcement, it remains to show, by other experiments, that the frustration effect operates independently of, and in addition to, other effects which have already been conceptualized.

Summary

This experiment was designed to test the assumptions that frustration is a motivational state, and that its strength is related to time in the frustrating situation.

Eighteen male albino rats were trained under hunger drive to run down an alley into a goal box, then leave that goal box and run down a second alley into a second goal box distinctively different from the first one. Their time to traverse the alley between Goal box 1 and Goal box 2 was measured during a preliminary period. When this time had reached a stable minimum, their running time and latency in leaving Goal box 1 were measured during test trials. In half of these trials they were frustrated in Goal box 1; in the other half, they were not. During this test period Ss were split into subgroups differentiated according to time spent in Goal box 1 on the frustration trials (5-, 10-, and 30-sec. groups).

An implication of the first assumption—that strength of performance on frustration test trials should be greater than that during the preliminary trials or that on reward test trials—was strongly supported by the data. There was, however, no statistically significant empirical support for the hypothesis of variation in strength of frustration when the time in the frustrating situation was varied between 5, 10, and 30 sec.

(Manuscript received July 23, 1951)

References

1. Amsel, A. The effect upon level of consummatory response of the addition of anxiety to a motivational complex. *J. exp. Psychol.*, 1950, **40**, 709–715.
2. Bruce, R. H. An experimental investigation of the thirst drive in rats with especial reference to the goal-gradient hypothesis. *J. gen. Psychol.*, 1937, **17**, 49–60.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Maltzman, I. The process need. *Psychol. Rev.*, 1952, **59**, 40–48.
5. Miller, N. E., & Stevenson, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.
6. Morgan, C. T., & Fields, P. E. The effect of variable preliminary feeding upon the rat's speed of locomotion. *J. comp. Psychol.*, 1938, **26**, 331–348.
7. Rohrer, J. H. A motivational state resulting from non-reward. *J. comp. physiol. Psychol.*, 1949, **42**, 476–485.

VISUAL PERCEPTUAL LATENCY AS A FUNCTION OF STIMULUS BRIGHTNESS AND CONTOUR SHAPE [1]

PAUL G. CHEATHAM

U. S. Navy Electronics Laboratory

Many studies on visual perception have been reported in which stimulus contour has been the independent variable (**1, 2, 3, 8, 9, 14, 15, 24, 25, 26, 27**). In none of these, however, has the latency of contour perception been investigated. In the present experiment, the problem was to measure perceptual latency as a function of variations in the stimulus contour shape and the stimulus brightness.

The results of the various studies on contour which are relevant to this experiment may be summarized as follows:

1. The stimulus correlate for the perception of contour is an abrupt change in stimulation between adjacent retinal areas, each of which must be approximately homogeneous, but must be different from the other in either color or brightness.

2. Contour is fundamentally significant for the visual perception of objects. Contour must be formed before the object can be perceived.

3. A certain time is required for the development of contour in perception, and if this time is not sufficient for complete development, the contour and the surface qualities within the contour will not be perceived.

4. The basic physical correlate for contour shape is the relative spatial distribution of the two bright or colored areas forming the gradient.

[1] Submitted in partial fulfillment of the requirements of the Ph.D. degree in the Division of Psychology, Department of Graduate Studies, University of Virginia. The writer is greatly indebted to Dr. Frank A. Geldard for his assistance, guidance, and encouragement throughout the course of this research.

The problem of measuring perceptual latency.—In this paper perceptual latency is defined as the interval of time which elapses from the beginning of the objective stimulus action on the receptor to the beginning of the related perception of the stimulus. This latency is exceedingly short. It must be measured independently of any motor reaction time. Because of these two factors, the technical difficulties of measurement are enormous.

Two techniques for the measurement of perceptual latency were found which could be adapted to this research, one developed by Fröhlich (**5**) for the measurement of sensation time (Empfindungszeit), the other by Piéron (**21**) for determination of visual latent periods. The Fröhlich method was not used here because of the doubt cast on the validity of its basic principles by the criticism of other writers (**4, 6, 7, 13, 17, 22, 23**). The Piéron method is called the masking method (méthode de masquage), and has been investigated independently by Monjé (**18**) under the name of the extinction method (Methode des Löschreizes). In this method, the latency of one stimulus is determined in relation to the constant latency of another stimulus, both within the same sense modality. Thus, only a relative latency is being measured which, however, should be just as valuable for study as an absolute latency, inasmuch as the same temporal interval is being studied, and the processes occurring during this interval should be the same.

A brief description of the technique will be of help in understanding the basic principles involved in perceptual latency measurement. Two stimuli are used which are presented to the same retinal area: (*a*) a small test stimulus of short duration and low brightness, and (*b*) a large masking stimulus of short duration and fixed high brightness. When the weak test stimulus is presented simultaneously with the strong masking stimulus, only the masking stimulus will be perceived. The test stimulus has been masked. If this brightness relationship is held constant, and an interval of time is introduced between the presentation of test stimulus and masking stimulus, a value of this interval may be found at which the two stimuli are perceived as being separate in time. The procedure here is to present the masking stimulus at progressively longer temporal

intervals after the presentation of the test stimulus. The maximum interval between the two for which the test stimulus is just masked by the masking stimulus is called the *limiting masking interval.* This interval may be determined accurately by proper psychophysical methods.

Piéron explains the masking phenomenon in terms of latency. The masking stimulus alone will be perceived as long as the interval of time separating the two stimuli is shorter than the latency of the test stimulus. The masking stimulus, with its very short latency, will be perceived first, even though presented objectively later than the test stimulus. The test stimulus will not be perceived at all if the masking stimulus is perceived first. If the interval exceeds the latency of the test stimulus, both stimuli will be seen. The test stimulus, in this case, will appear first and the masking stimulus second, or the two will appear simultaneously, the test stimulus being superposed upon the masking stimulus. The difference in their latencies has been compensated for by the physical temporal interval introduced between their objective presentations. Two assumptions are needed for this explanation: (*a*) the latency of a visual response is inversely related to the intensity of the stimulus, and (*b*) the latency of the very bright masking stimulus is of a minimal value, that is, its latency is as short as possible. With these assumptions, the limiting masking interval may be regarded as a direct measure of the latency of the test stimulus, relative to the constant short latency of the masking stimulus.

Neither Piéron nor Monjé studied latencies in relation to variations in stimulus contour. Piéron obtained functions relating latency to the variables of intensity, duration, area, retinal location, and wave length of the test stimulus. Intensity was the only variable in the Monjé experiment.

The theory of perceptual latency.— The explanation given by Piéron for the peripheral events of masking does not take into account the central processes which occur after the neural impulses reach the cortex. He left this central phase unexplained, saying that when a strong stimulus masks a weaker one, it is because the latent time of the stronger stimulus is shorter, and its perceptive reaction has been produced before that of the weaker one. The question to be asked next is this: Why are the neural impulses set off by the weak test stimulus not able to produce a perceptive reaction once the impulses of the masking stimulus have reached the cortical centers and established their reaction first? A process has occurred which cannot be explained by a peripheral theory alone. Some type of cortical field theory seems to be needed. The satiation theory presented by Köhler and Wallach (16) would seem to offer a satisfactory framework into which the facts of masking can be set.

According to the satiation theory, inspection of contours produces electrochemical figure currents which flow through the cortical tissue. The process of satiation is the electrotonic effect of figure currents. The degree of satiation produced at any time during inspection is a function of the intensity of the figure currents, and this, in turn, is proportional to the difference in the frequency of neural discharge from the receptor, and thus to the stimulus correlate of the brightness difference between two or more areas in the visual field. The distribution of satiation in the tissue follows that of the satiation figure currents; where currents are dense, satiation will develop more quickly and to a higher level than in other parts. Since figure currents have their highest density around the contour of the stimulus, satiation in these areas will have a higher intensity. Satiation is persistent, thus the alteration of subsequent currents entering the satiated field is demonstrable for an appreciable time (11).

One more characteristic of satiation must be discussed before the theory is ready to be applied to the masking phenomenon. Satiation is a process, which means that changes in satiation must be a function of time. Köhler and Wallach do not make clear just what direction the change takes at the very beginning of the process. They do indicate in several places (16, p. 348 and 354) that satiation would increase initially before any limiting process would set in, and that an inspection period of as little as .5 sec. should produce a certain degree of satiation. It would follow that alterations in currents of a subsequent figure should be dependent upon the moment during the satiation process at which the figure currents enter the satiated field.

One may hypothesize that the masking stimulus produces figure

currents, since the difference in brightness between stimulus and ground is sufficient to establish such currents. The level of satiation created by these currents should build up quickly and should reach a fairly high degree during the action of the stimulus. This level of satiation should persist after the stimulus has been removed. Any subsequent stimulus, such as the test stimulus, acting in this region would produce currents of its own which would be affected by the level of satiation of the masking stimulus. The point in time, and thus the level of satiation reached, at which the test stimulus enters the satiated field of the masking stimulus would determine the effectiveness of the test stimulus. The brightness of the test field would be a determining factor also, since figure currents are directly related to brightness. If the test stimulus enters the satiated area after an interval of time sufficiently long for a high degree of satiation to have been reached, and the brightness of the test stimulus is low, producing figure currents of low intensity, then the test stimulus will not be effective. In this case, the test stimulus has not been able to "break through" the satiated field of the masking stimulus. If, on the other hand, the test stimulus enters after a short interval of time, the satiation process of the masking field having just begun, it would be effective. This is precisely what occurs in the masking phenomenon. It must be remembered, however, that the interval elapsing before the test stimulus enters is inversely related to the objective presentation of the two stimuli. Thus, when the masking stimulus occurs at greater and greater objective intervals after the test stimulus, the physiological interval between their latencies becomes smaller and smaller.

One more assumption is needed before the explanation is complete. The hypothesis may be given that some sort of "breaking through" process would take place when figure currents of a test stimulus are produced in a previously satiated field. This process might be analogous to the spark gap phenomenon. When satiation of the masking field is of a given degree, any other stimulus entering that field must be of sufficient intensity in relation to its ground so that its currents can break through into the field. The currents related to a given figure may become effective within a given satiated field by two means: (*a*) the intensity of the figure may be increased, or (*b*) the figure may be made to enter the field before the process of satiation has reached a high level. The latter means is used in the masking technique. It follows that the time required for a perception to develop, which is measured by the masking technique, is made up not only of physiological latency of impulses traveling from peripheral to central areas, but also of the time required for the breaking through process.

Hypotheses.—Two hypotheses have been drawn from the preceding theory for experimental test:

I. For any given distribution of two adjacent homogeneous surfaces sufficiently different in brightness to form a contour, the perceptual latency is a function of the degree of this brightness difference; the greater the difference in brightness, the shorter the perceptual latency.

II. For any given brightness difference, the perceptual latency is a function of the distribution of surface brightness forming the contour, when each surface remains homogeneous throughout and constant in total area.

Hypothesis I states that perceptual latency should be inversely related to stimulus brightness. An increase in test stimulus brightness would produce a decrease in physiological latency and an increase in the figure current intensity. Consequently the test stimulus would be able to break through a satiated field of higher degree.

Hypothesis II states that different contours should require different times to become effective. The strongest evidence that this hypothesis might be confirmed comes from studies on contour by Werner (**24**, **25**). He produced experimental evidence that angular parts of contours have more "psychophysical energy" than circular parts, and that angles are points of concentrated energy in the contour. Angular contours, then, might have shorter perceptual latencies than non-angular ones. Support for this hypothesis may be found in the satiation theory. Currents associated with the angular figure would have their greatest density nearest the contour. Within a given angle of the figure, the contours are close together, therefore the current would be more dense within the angle than elsewhere in the figure. Angles, then, should be more capable of breaking through a previously satiated field than circular figures, when the two are equal in all other physical variables which are known to affect latency. Any difference found in their perceptual latencies would be due to this superior breaking through ability of the angular figure, and this ability must be due solely to the presence of the angles. The physiological latency, determined by duration, area, and brightness, would be the same for both figures.

Method

Apparatus.—The apparatus was designed to satisfy three experimental needs: (*a*) independent manipulation of the temporal interval between the test and masking stimuli, preserving at the same time the constant durations of both stimuli; (*b*) variations in the contour of the test stimulus independently of the contour of the masking stimulus; (*c*) variations in the brightness of the test stimulus independently of the constant brightness of the masking stimulus.

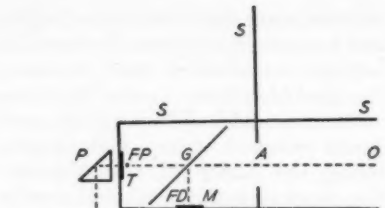

FIG. 1. Schematic diagram of the apparatus

A schematic diagram of the apparatus is given in Fig. 1. The source of light for the test stimulus was a 500-w. projector lamp (*LP*) operating on a constant voltage of 110 v. DC. A large-diameter condensing lens (*L1*) brought the beam of light to a sharp focus in the plane of the exposure disk (*D*). The beam diverging beyond the focal point was made parallel by a double condenser lens system (*L2*). Placed in the path of this beam were the necessary devices for controlling the intensity and quality of the light, which were a Wratten neutral tint wedge and balance (*W*), Wratten neutral tint percentage and decimal filters (*FN*), and a Wratten No. 78 daylight filter (*FD*). The beam was bent through a 90° angle by a large totally-reflecting prism (*P*) and then directed onto the flashed opal glass plate which was held in a frame fixed to an opaque screen. The opal glass formed a homogeneously diffused surface of brightness for the test field. The test contour (*T*) was a black cardboard cutout slide which was placed in the frame directly in front of the opal glass. The area of the cutout part was thus made homogeneous in brightness, whereas the remainder of the field was uniformly dark. The maximum brightness of the test field was 4.6 ml.

A 750-w. lamp provided the light source for the masking stimulus. The beam was made parallel by a double condenser system (*L3*), and was then converged by another lens (*L4*) to a sharp focus in the plane of the disk. Diverging from this focus, the beam fell on an opal glass plate which served as the diffusion surface for the masking stimulus (*M*). A daylight filter (*FD*) was placed directly in front of the opal glass. A cardboard cutout formed the contour of the masking stimulus and was inserted into the frame in front of the daylight filter. The brightness of the masking field was 7.0 ml.

A large plate of window glass (*G*) was placed at a 45° angle to the masking and test fields. The test stimulus was visible to *O* by direct transmission through the glass, whereas the masking stimulus was seen by reflection off the surface of the glass. In this manner the two stimuli were made to fall on the same region of the retina, yet each could be varied independently with respect to contour, brightness, and position.

A small red fixation point (*FP*) was placed tangent to the top of the test stimulus contour, and was visible to *O* at all times. Opaque screens (*S*) were placed around the stimuli and *O* to eliminate all stray light. The entire inside of the apparatus was painted flat black to form a homogeneous dark ground for the stimuli.

The exposure disk was of the Whipple type. Two apertures 180° apart cut the focal points of lenses *L1* and *L4*, giving instantaneous exposure and termination of the two lights. Duration of exposure was set at 100 msec. for both stimuli. The angular separation of the apertures, and thus the temporal interval between exposures, could be varied by changing the position of the aperture exposing the masking stimulus. This interval was calibrated in milliseconds with a range of variation from zero to 600 msec. Average error of calibration, taken over a period of several days, was less than 1 msec., as measured by a photoelectronic counter of high accuracy.

The masking stimulus was a circle 30 mm. in diameter and 706 sq. mm. in area. It subtended a visual angle of 1°26′. The test stimuli contours were a circle (15 mm. diameter, 44′ subtense), a triangle (15 mm. altitude, 23.5 mm. base, 44′ subtense for altitude), a square (13 mm. side, 38′ subtense), and a diamond, which was the square set on one angle. All test figures had the same approximate area of 176 sq. mm. Observation distance used in computing subtenses was 120 cm. The center of the test stimulus always coincided with the center of the masking stimulus. While the test figures were equal in total area, they were not the same shape, consequently they did not fall on the same retinal surface throughout. The slight difference in retinal receptors activated by nonoverlapping parts of the figures was not considered to be important since the figures fell completely within the fovea.

Procedure.—Initial dark adaptation of *O* lasted about 15 min. During this time he was given instruction as to the nature of the stimulus and the method of reporting. The word "Ready" was spoken by *E* as the ready signal about 2 sec. before the stimuli were presented. The *O* reported immediately after, describing in detail exactly what he observed with respect to positional changes, brightness changes, and changes in figure contour. He reported "Positive" when he saw a complete, distinct contour. He reported "Negative" when he saw only a homogeneous field of the masking stimulus. Rest periods of 1 min. were given between exposures. This was sufficient to eliminate after-images.

The method of limits provided an accurate measure of the limiting masking interval. Ascending series were begun with an interval giving always a negative report. The interval was increased progressively until *O* reported the first appearance of any contour. Three successive positive reports were demanded beyond this point of change. Descending series were similar, beginning with an interval giving a positive report, and ending with three successive negative reports. The ascending and descending series were presented in varying lengths, and were alternated in counterbalanced order. Short rest periods between groups of six series were given.

The contours were presented in random order for each brightness level. A different contour was used for each new session. The brightness was changed for each cycle of contours. A session consisted of 15 min. of dark adapting and 1 hr. of data taking.

Results

Preliminary trials showed that there was a fairly large zone of transition from complete homogeneity to complete contour. Within this zone, the contour was constantly changing in degree of completeness, position of incomplete parts, and brightness. This zone of transition was not predicted on the basis of the Piéron (**19, 20, 21**) and Monjé (**18**) experiments. In a typical ascending series, there were three major changes in the

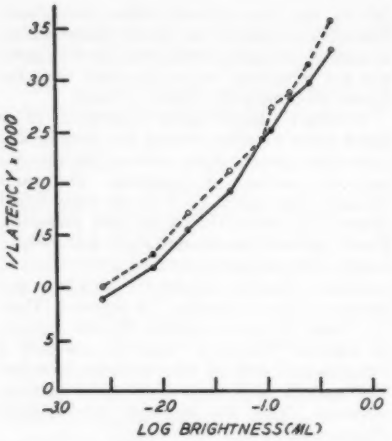

FIG. 2. The reciprocal of perceptual latency (broken line) and development time (solid line) as a function of the logarithm of the stimulus brightness

developmental process from homogeneity to complete contour. The first was the break in the homogeneity of the masking field, observed as a brightening of the field in the center. This brightness remained for several millisecond samplings before the next change occurred. The second change occurred when the brightened portion of the field developed into some part of the stimulus contour. Only one side or one angle might appear. The contour continued to develop, becoming more and more complete until the whole contour was finally perceived. Once the complete contour was perceived, it remained complete with further sampling of the process. This sequence was reversed for the descending series.

The appearance of a developmental sequence of this type required a definition for the two points of transition. The lower transition point was defined as the change from brightening to first contour, and the time measured here was regarded as representing the perceptual latency for contour, since this first appearance of contour terminated a period of time during which there was no objectively observable activity. The upper transition point was defined as the change from incomplete to complete contour. The time at this point was called the development time for contour, and represents the time required for the contour to become developed completely.

Tests of hypotheses.—Hypothesis I predicted an inverse relationship for perceptual latency and test-figure brightness. The curve in Fig. 2 shows that this prediction was approximately correct. Each point on the curve represents the mean of 24

TABLE 1

PERCEPTUAL LATENCY AS A FUNCTION OF CONTOUR SHAPE

Brightness in ml.	O	Contour Shape							
		Circle		Square		Diamond		Triangle	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
.46	D	17.1	3.2	19.8	3.0	22.3	2.4	20.3	4.1
	T	27.8	1.5	28.4	2.2	31.6	1.6	30.5	1.4
.046	D	34.3	5.2	34.6	3.0	31.0	3.4	23.6	4.6
	T	40.1	2.2	40.4	2.5	44.0	2.3	42.3	1.3
.0046	T	79.2	3.4	75.6	2.8	85.6	5.1	79.1	3.0

TABLE 2

DIFFERENCES AND *t* VALUES FOR DIFFERENCES IN PERCEPTUAL LATENCY BETWEEN CONTOURS

Brightness in ml.	*O*	Contours Compared											
		Circle-Square		Circle-Diamond		Circle-Triangle		Square-Diamond		Square-Triangle		Diamond-Triangle	
		Diff.*	*t*	Diff.	*t*	Diff.	*t*	Diff.	*t*	Diff.	*t*	Diff.	*t*
.46	D	−2.7	2.04	−5.2	*4.33***	−3.2	2.04	− 2.5	*2.19*	−0.5	0.33	2.0	1.41
	T	−0.6	0.75	−3.8	*5.75*	−2.7	*4.50*	− 3.2	*3.95*	−2.1	*2.72*	1.1	1.74
.046	D	−0.3	1.65	3.3	1.76	10.7	*5.09*	3.6	*2.68*	11.0	*6.66*	7.4	*4.33*
	T	−0.3	0.30	−3.9	*4.10*	−2.2	*2.89*	− 3.6	*3.52*	−1.9	*2.26*	1.7	*2.12*
.0046	T	3.6	*2.70*	−6.4	*3.44*	0.1	0.73	−10.0	*5.64*	−3.5	*2.82*	6.5	*3.61*

* A negative difference indicates that the second contour compared had the longer latency.
** Italicized *t* values indicate significance at the 5% level or better.

series, 12 ascending and 12 descending. Perceptual latency approaches a straight-line function similar to the conventionally accepted inverse relationship of latency and intensity found by electrophysiological methods (**10, 12**). The variability around the separate means was extremely small, having an average magnitude of around 2.3 msec., and a range of 10. to 4.4 msec.

The data for the test of Hypothesis II are given in Table 1. Each mean was obtained from 12 ascending and 12 descending series.

The *t* test was applied to each pair of contours, with each contour being compared with every other contour, to determine the significance of difference between mean perceptual latencies. This test assumes a normal distribution, but latency measures are usually somewhat skewed. Therefore, another test of significance, the Sign Test (**28**), which does not assume a normal distribution, was applied to the data. This test agreed in every case with the *t* test, thus the data for the latter have been reported. Table 2 contains the mean differences in perceptual latency, the *t* ratios, and the level of significance for all combinations of contours at three brightness levels.

Table 2 shows several facts: (*a*) 20 of the 30 comparisons were significant at the 5% level; (*b*) 6 of the 12 comparisons at .46 ml., 9 of the 12 at .046 ml., and 5 of the 6 at .0046 ml. were significant at the 5% level. The data suggest that perceptual latency might be a function of contour shape with further sampling. One must realize that the judgment at this transition point was extremely difficult since it required *O* to report the first appearance of any contour whatsoever. Consequently, these latencies were not very reliable, thus making less significant any small differences which might have appeared with changes in contour.

In Fig. 3 are shown the curves resulting from a plot of mean perceptual latencies as a function of contour shape. The contours were arranged in order according to the hypothesis that latency is a function of the angularity of the contour, and that the smaller the angles, the shorter the latency. If this were true, the curves would show a rise in latency from

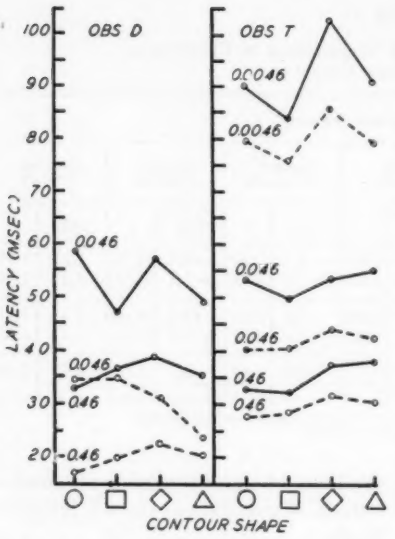

FIG. 3. Perceptual latency (broken lines) and development time (solid lines) as a function of contour shape, with the brightness level as a parameter

triangle to circle. It is immediately apparent that this trend was not present. Marked individual trends were present, however.

Comparisons by the *t* test showed that there is some difference between the two groups of data which is not due to sampling variation, but is attributable to the independent variable. Since distribution of contour was the only variable manipulated, the inference can be made that the difference is due to this variable. No inference can be made, however, of the direction of difference, since differences of opposite sign but of the same absolute magnitude would also be significant by the *t* test. However, if all of the differences for one condition of measurement have the same sign, and all are significant, then one could accept the differences and their sign as indicative of the true ones.

The development time for contour differed significantly in general as a function of contour shape. The data are given in Table 3 and Table 4. Application of the *t* test showed the following facts: (*a*) 20 differences out of 30 tested were significant at the 1% level; (*b*) 8 of the 12 at .46 ml., 7 of the 12 at .046 ml., and 5 of the 6 at .0046 ml. were significant at the 1% level. Figure 2 shows that development time varies inversely with stimulus brightness, and Fig. 3 shows that development time varies with contour shape, with these variations following trends similar to those for perceptual latency.

The trends for perceptual latency and development time were only

TABLE 3

DEVELOPMENT TIME AS A FUNCTION OF CONTOUR SHAPE

Brightness in ml.	O	Contour Shape							
		Circle		Square		Diamond		Triangle	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
.46	D	32.8	.2.1	36.5	1.1	38.6	1.7	35.2	2.6
	T	32.8	1.6	32.2	2.3	37.4	0.9	37.9	2.3
.046	D	58.7	4.0	47.0	2.4	57.0	2.7	49.0	2.6
	T	53.1	2.6	49.7	2.9	53.5	1.4	55.0	2.0
.0046	T	89.9	1.8	83.7	3.0	102.3	3.8	90.4	2.6

TABLE 4

DIFFERENCES AND *t* VALUES FOR DIFFERENCES IN DEVELOPMENT TIME BETWEEN CONTOURS

Brightness in ml.	*O*	Contours Compared											
		Circle-Square		Circle-Diamond		Circle-Triangle		Square-Diamond		Square-Triangle		Diamond-Triangle	
		Diff.*	*t*	Diff.	*t*	Diff.	*t*	Diff.	*t*	Diff.	*t*	Diff.	*t*
.46	D	−3.7	*5.28***	− 5.8	*7.25*	−2.4	2.03	− 2.1	*3.50*	1.2	1.35	3.4	*3.50*
	T	0.6	0.72	− 4.6	*8.52*	−5.1	*6.14*	− 5.2	*7.12*	−5.7	*5.93*	−0.5	0.68
.046	D	11.7	*8.29*	1.7	1.17	9.7	*6.73*	−10.0	*9.34*	−2.0	1.88	8.0	*7.14*
	T	3.4	*2.93*	− 0.4	0.46	−1.9	1.93	− 3.8	*4.00*	−5.3	*5.00*	−1.5	2.02
.0046	T	6.1	*5.75*	−12.5	*9.84*	−0.6	0.63	−18.6	*12.74*	−6.7	*5.63*	11.9	*8.62*

* A negative value indicates that the second contour compared had the longer latency.
** Italicized *t* values indicate significance at the 1% level or better.

suggestive of those which might be shown to exist with further sampling. The obtained trends did not agree with the hypothesized trends, and were, in fact, almost the reverse. In general, the circle and square were perceived more rapidly than the triangle and diamond. Of interest is the fact that all square-diamond differences were significant. The diamond was formed by placing the square on one angle, keeping all other physical properties of the two figures identical.

Qualitative observations.—The following changes of a qualitative nature were reported by *O*s: (*a*) The masking field became brighter on both ascending and descending series just before the first or just after the last appearance of contour. This result would indicate that a brightness gradient greater than one j.n.d. was reached before contour was formed, a fact which is significant in relation to the definition of contour given earlier. The latency for light appeared to be shorter than the latency for contour. (*b*) The developmental sequence showed that sides of the figure appeared first, and a good portion of two adjacent sides must have been formed before the angle between developed. In many instances, all sides of the figure were developed completely before the first angle appeared. In the case of the circle, portions of the contour appeared randomly in any position. (*c*) During a set of investigations subsequent to the main experiment, hue was introduced into the contour. The hue never became perceived until the complete contour was formed. That is, during the developmental sequence of the contour, *O* reported no hue even though hue was introduced physically into the light path. This result was surprising and would seem to indicate that the latency for hue is longer than the latency for either light or incomplete contour. In a typical ascending sequence, the homogeneous field appeared colorless, then brightened, then developed into portions of incomplete contour, and finally appeared colored when the complete contour was formed. The *O* was sometimes surprised to find hue in the complete contour when he had not perceived hue up to that point.

Discussion

The results showed that perceptual latency and development time for contour, as determined by a masking method, vary inversely with the test stimulus brightness in a manner similar to functions obtained by electrophysiological methods. The masking method, then, apparently gives values of perceptual latency which are measurable in the human observer independently of any motor response, and which are quantitatively precise enough to warrant their use for further investigations.

Extension of the experiments reported here seems necessary, inasmuch as the hypothesis concerning angularity of contours was not found to be true under the conditions of the experimental test made of it. Angular contours had longer latencies rather than the hypothesized shorter latencies. Another hypothesis might be suggested to take care of this result. The perceptual latency of any contour should be a function of the proximity of the test stimulus contour to the masking stimulus contour, increasing as this distance decreases. Evidence that contours affect the brightness threshold of adjacent contours may be found in the experiments of Fry and Bartley (1, 8). The angular parts of the square, diamond, and triangle were closer to the contour of the masking field than was the perimeter of the circle. These figures not only developed with a longer latency, but the angles themselves were the last part of the contour to develop. This hypothesis could be drawn from the satiation theory also. Figure currents are more dense around the contour, thus in the masking circle satiation would develop more rapidly and reach a higher level around the contour than in the center. A test figure would be less effective in those regions near the contour of the masking stimulus where satiation is higher. A brief preliminary test of this hypothesis was made using the square and the diamond at two brightness levels. The data gave a fair indication that the position of the test stimulus contour in relation to the masking stimulus contour would produce changes in the perceptual latency. The differences in latency between the peripheral and central positions were not all significant, but the latencies for the peripheral positions were consistently longer than those for the central positions. Further tests of this hypothesis are needed.

Perhaps the primary determinant is not angularity or proximity of contour, but is some related but different factor like compactness. Other more conceptual factors such as "goodness" or "articulation" might be important, but these would have to be reduced to some variable aspect of the physical stimulus before they could be introduced as independent variables.

The problem of localization presents itself. The assumption was made that all processes and events occurring in relation to perceptual latency and masking were central in nature. Central, in this case, would mean cortical. The binocular observation used, however, would not give any indication that the process was cortical and might not have taken place entirely within the retinal levels. Köhler and Wallach (16) have emphasized the central location of the satiation process. The level of localization for masking could be determined by presenting the masking stimulus to one eye and the test stimulus to the other, combining the two by stereoscopy. If masking takes place under these conditions, one could say that the process is centrally located.

The possibility of retinal field action is not to be ignored. Interaction effects of both inhibitory and facilitative nature have been demonstrated in the retina. However, the present information about the physiology and anatomy of retinal elements, and the ways in which these elements work together, does not fit well into a field theory. On the other hand, it does not preclude a retinal field theory.

The question of spatial interaction might be approached using the measure of latency instead of the conventional measures of CFF, absolute brightness threshold, or differential brightness threshold. Spatial interaction was indicated in connection with the effects of test stimulus proximity to masking stimulus contour. Basic stimulus arrangements could be introduced into the masking situation, and the latency of one of the test fields could be measured as it is affected by the presence of the other field. Perhaps there might be a definite alteration in the temporal processes associated with spatial interaction.

Individual differences were marked. A study of perceptual latency for a large sample of *Os* should be of value. One might hypothesize that the time required to perceive is related to certain physical and psychological characteristics of the individual. Such relationships involving a trait as basic as perceptual latency would be extremely significant.

The question of theoretical importance which should be asked is this: What is perceptual

latency? In terms of present neurophysiological knowledge, what physiological process takes place which requires time, when this time is a function of variation in the stimulus contour? The only answer at present is theoretical. The processes requiring time which are a function of stimulus contour are the building up process of satiation and the breaking through process of the figure currents, both hypothesized in this experiment.

Summary

In these studies perceptual latency and perceptual development time were measured as a function of changes in the brightness and the contour shape of the stimulus. Measurement was by the method of masking, which gave values of relative perceptual latency, obtained in *O* independently of any motor reaction time.

An inverse relationship was found between the test stimulus brightness and the perceptual latency for that stimulus.

Perceptual latencies did not, in general, vary significantly with variations in the stimulus contour. The data did not support the hypothesis that the more angular figures would have shorter latencies. Actually, angles and angular figures had the longest latencies.

Development time, measured as that time required for the complete development of the contour, was inversely related to stimulus brightness. Unlike latency, it did vary significantly with changes in contour shape.

During the developmental period, the contour passed through an orderly sequence. A small portion appeared first, then increased in size, and finally became complete. Sides appeared before angles.

(Manuscript received August 27, 1951)

References

1. Bartley, S. H. *Vision.* New York: D. van Nostrand, 1941.
2. Collier, R. M. An experimental study of form perception in indirect vision. *J. comp. Psychol.*, 1931, **11**, 281–289.
3. Douglas, A. G. A tachistoscopic study of the order of emergence in the process of perception. *Psychol. Monogr.*, 1947, **61** (6), Whole No. 287.
4. Erismann, T. Die Empfindungszeit; eine kritische Untersuchung. *Arch. ges. Psychol.*, 1935, **93**, 453–519.
5. Fröhlich, F. W. *Die Empfindungszeit. Ein Beitrag zur Lehre von der Zeit-, Raum- und Bewegungsempfindung.* Jena: G. Fischer, 1929.
6. Fröhlich, F. W. Zur Kritik der Empfindungszeitmessung. *Z. Psychol.*, 1931, **121**, 357–363.
7. Fröhlich, F. W. Bemerkungen zu G. E. Müllers Kritik der Empfindungszeitmessung. *Z. Sinnesphysiol.*, 1932, **62**, 246–249.
8. Fry, G. H., & Bartley, S. H. The effect of one border in the visual field upon the threshold of another. *Amer. J. Physiol.*, 1935, **112**, 414–421.
9. Geissler, L. R. Form perception in indirect vision. *Psychol. Bull.*, 1926, **23**, 135–136.
10. Graham, C. H. Vision: III. Some neural correlations. In C. Murchison (Ed.), *A handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934.
11. Hammer, E. R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, **62**, 337–354.
12. Hartline, H. K. Neural mechanisms of vision. *Harvey Lect.*, 1941–1942, **37**, 39–68.
13. Hecht, S. Die Empfindungszeit. *Naturwissenschaften*, 1930, **18**, 834–835.
14. Helson, H., & Fehrer, E. V. The role of form in perception. *Amer. J. Psychol.*, 1932, **44**, 79–102.
15. Hochberg, J. E., Gleitman, H., & MacBride, P. D. Visual threshold as a function of simplicity of form. *Amer. Psychol.*, 1948, **3**, 341–342.
16. Köhler, W., & Wallach, H. Figural after-effects. An investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
17. Metzger, W. Versuch einer gemeinsamen Theorie der Phänomene Fröhlichs und Hazelhoffs und Kritik ihrer Verfahren zur Messung der Empfindungszeit. *Psychol. Forsch.*, 1932, **16**, 176–200.
18. Monjé, M. Die Empfindungszeitmessung mit der Methode des Löschreizes. *Z. Biol.*, 1927, **87**, 23–40.

19. Piéron, H. De la variation des intervalles limites de masquage d'une excitation lumineuse par une excitation consécutive très intense en fonction de l'intensité de la première. *C. R. Soc. Biol. Paris*, 1923, **88**, 736–789.
20. Piéron, H. Détermination du retard de l'excitation nerveuse par les produits de la réaction photo-chemique rétinienne moyen d'une méthode de mesure des intervalles limites de masquage d'une excitation luminuese variable par une excitation très intense. *C. R. Soc. Biol. Paris*, 1923, **88**, 689–692.
21. Piéron, H. Recherches expérimentales sur la marge de variation du temps de latence de la sensation lumineuse (par une méthode de masquage). *Année psychol.*, 1925, **26**, 1–30.
22. Rubin, E. Kritisches und Experimentelles zur "Empfindungszeit" Fröhlichs. *Psychol. Forsch.*, 1929, **13**, 101–112.
23. Rubin, E. Bemerkungen zur Ewiderung Fröhlichs. *Psychol. Forsch.*, 1930, **13**, 289–292.
24. Werner, H. Studies on contour: I. Qualitative analyses. *Amer. J. Psychol.*, 1935, **47**, 40–64.
25. Werner, H. Strobostereoscopic phenomena. *Amer. J. Psychol.*, 1940, **43**, 418–422.
26. Wever, E. G. Figure and ground in the visual perception of form. *Amer. J. Psychol.*, 1927, **38**, 194–226.
27. Wilcox, W. Helson and Fehrer on the role of form in perception. *Amer. J. Psychol.*, 1932, **44**, 578–580.
28. *The statistical sign test.* Princeton University Research Group. OSRD 1945. Publ. Bd. No. 23726. U. S. Dept. Commerce, Washington, D. C., 1946.

PERFORMANCE DECREMENT IN A SIMPLE MOTOR TASK BEFORE AND AFTER A SINGLE REST[1]

EDWARD A. BILODEAU[2]

State University of Iowa

Almost all recent research on motor behavior has dealt with tasks with which *S* gradually acquires considerable proficiency. Tasks such as the pursuit rotor, inverted alphabet printing, and the Mashburn have been most frequently used. In the theoretical analysis of the data, two different intervening variables have been postulated; one for the learning factor ($_sH_R$) and one for the work factor (I_R). $_sH_R$ and I_R usually have varied together and have been studied in controlled situations where the effects of one can sometimes be isolated from those of the other (6).

It is equally desirable to isolate more directly the effects of work from those of learning. This can be accomplished by devising a situation where the effect of one remains constant and equal among conditions throughout practice. Studies with the ergograph (8) have these characteristics, but unfortunately the experimental designs have not been such as to facilitate the interpretation of motor behavior in the light of more recent developments.

In the present study a new device which presumably allows for little learning has been constructed. Decrements in performance have been induced through manipulations of the amount of work required to accomplish a relatively simple movement as well as through the more customary procedure of varying the length of the practice period. It was thus hoped that by systematic variation of the conditions of work in a situation where little or no learning takes place the effects of work could be isolated from those of learning.

The purpose of the present study was to investigate several aspects of response decrement and recovery from decrement as functions of three independent experimental variables: (*a*) the amount of work required to accomplish a relatively simple movement (work-loading); (*b*) the duration of a single initial practice period (initial practice); and (*c*) the duration of a single rest period following the initial practice period (rest).

These variables are operative in any motor task whether simple or complex. All have been independently studied in both ergographic and pursuit tasks. No experimenter, however, using a nonlearning type of task has studied response decrement and recovery in terms of their functional relationships to one of the three variables with the remaining two systematically controlled.

METHOD

Subjects.—The *S*s were 256 women obtained from the undergraduate classes in psychology at the State University of Iowa. They were run in pairs and encouraged to compete or race with each other.

Apparatus.—The major elements of the apparatus were two identical wooden chutes (17

[1] This article represents part of a dissertation submitted to the Graduate School of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Dr. Judson S. Brown, under whose direction this research was conducted.

[2] Now at Perceptual and Motor Skills Research Laboratory, Air Training Command Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

in. high and 1.5 in. square) and 32 lead-weighted table-tennis balls. The chutes were mounted about 3 ft. apart on a table top. A table-tennis ball lifted from a cup at the base of the chute and dropped into the top of the chute returned to the cup where it was again in a position ready for lifting.

The 32 balls were divided by weight into four sets of eight balls each. The balls in the four sets weighed 21, 42, 84, and 168 gm., respectively. Each pair of *S*s used the eight balls of one set, i.e., four balls for each *S*. Four balls assured a continuous flow through the chute and a constant supply at the cup for even the fastest *S*.

One *S* sat in front of each chute. Two counters were placed between the chutes so that they could be seen at a glance by either *S*. Whenever a ball was dropped into a chute an electromagnetic counter recorded the passage of the ball (a score of one). Two additional counters were placed in front of *E* and were read every 30 sec. Passage of a ball through a chute also activated one of two recording pens attached to a constant-speed polygraph, while a third pen produced a time line. The polygraphic record made it possible to score the number of balls lifted and dropped through either chute during *any* time interval desired.

Procedure.—The basic outline of the procedure consisted of an initial practice period of variable duration, a variable rest interval and a 4-min. period of final practice. The three-variable, $4 \times 4 \times 4$ factorial design employed is summarized in Table 1.

The *S*s were instructed to lift the weighted balls one at a time from the cup at the base of the chute and insert them into the top of the chute *as quickly as possible*. The instructions were such as to encourage *S*s to compete with one another and to use the counters to evaluate their relative performance levels. During the rest period the counters were reset to zero and *S*s were allowed to converse with each other and with *E*. No questions pertaining to the experiment were answered, and no instructions of any kind were given during the rest period. No information at all was provided about either the lengths of the forthcoming practice and rest periods or the weight of the ball to be lifted.

TABLE 1

THE THREE-VARIABLE FACTORIAL DESIGN

Variable	Values
Duration of initial practice	2, 4, 8, or 16 min.
Duration of interpolated rest	1, 2, 4, or 8 min.
Work-loading (ball weight)	21, 42, 84, or 168 gm.

The *S*s were randomly assigned by pairs to the 64 experimental conditions; after these *S*s were run the experiment was replicated by randomly assigning another pair to each condition. Since the number of *S*s within a condition (cell) was small, no attempt has been made to deal with the separate conditions. The data have been analyzed in terms of systematic combinations of conditions.

RESULTS

The results are presented in terms of the number of balls inserted into the top of the chute per 30 sec. Occasionally, however, successive pairs of 30-sec. scores were averaged in order to increase the stability of the measure.

A number of combinations are possible with the 64 factorial conditions employed and these are designated by a special notation; for example, 2-4-21 indicates progressively, initial practice (2 min.), interpolated rest (4 min.) and work-load (21 gm.). The symbol *X* is used to signify that a variate is a parameter (taking any one of the four experimental values), and the symbol *K* that an operation of summation over a particular set of four values has been made. Thus, *K*-*X*-*K* refers to that combination of conditions in which the entries for all durations of initial practice have been summed, in which rest is the parameter, and in which the entries for all values of work-loading have been summed. In this case *X* (rest) may take the values 1, 2, 4, and 8 min., and *K* represents summation over four different values of initial practice and over four values for work-loading.

Trend of the initial practice curve.—Since the performance of the 16-min. practice conditions is representative of the results obtained with 2, 4, and 8 min. of practice, the latter are not presented. The combination 16-*K*-*K* ($N = 64$), based upon all conditions

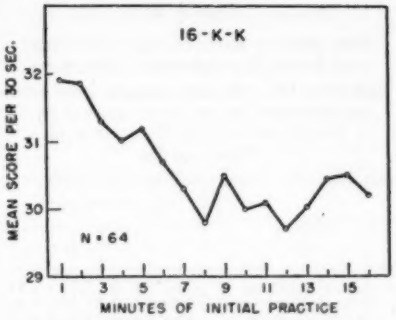

FIG. 1. Mean rate of responding during initial practice for the 16 conditions with 16 min. of practice

with 16 min. of initial practice, is presented in Fig. 1. There is no increase in score after 1 min. of practice, indicating that if there was any warm-up it must have taken place before the first minute of practice.[3] A decrease in rate of responding appears throughout the first 8 min. of practice.

Rate of responding during initial practice for each of the four loads is presented in Fig. 2 (successive pairs of 30-sec. scores were averaged for the pooled data of 8-*K*-*X* and 16-*K*-*X* for the first 8 min.). It is to be noted that the two combinations with the most extreme work-loads are most divergent while the means of the two remaining combinations (loads 42 and 84) fall fairly consistently below the means of the 21-gm. load.

For the first minute of practice, there is an inverse relationship between score and ball weight. Apparently the difference in mean rate of responding develops very quickly, for the difference among load means is significant beyond the 1% level of confidence when the first 30-sec. means are used. These values are 32.97, 32.65, 31.39, and 30.04 for loads 21, 42, 84, and 168, respectively. The standard error of the mean difference was calculated as .457, indicating that a difference between means must equal or exceed .90 in order to be significant at the 5% level of confidence.[4]

In order to ascertain whether differential rates of responding were obtained even prior to 30 sec. of work, polygraphic records of the first 5 sec. of initial practice were examined for random subsamples of 24 *S*s each from the 21-gm. and 168-gm. conditions. The mean for the 21-gm. sample was 5.70 compared with a mean of 5.08 for the 168-gm. sample ($\sigma_{diff.}$ equals

[3] The scores of Fig. 1 were obtained by averaging successive *pairs* of 30-sec. scores. Examination of the *individual* 30-sec. scores reveals some slight increase in rate of responding for 90 sec. The first four consecutive means were 31.70, 32.12, 32.17, and 31.59. A *t* for related measures was calculated between the first 30-sec. mean (31.70) and the highest 30-sec. mean (32.17) obtained during the entire 16 min. of practice. The hypothesis that these means were drawn from the same population can be rejected at the 7.5% level of confidence.

[4] The above analysis gives a general idea of the order of magnitude of differences required to meet statistical significance between 30-sec. means in later combinations.

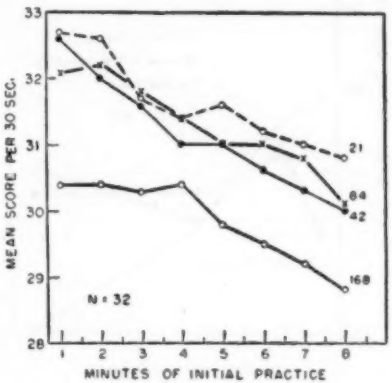

FIG. 2. Mean rate of responding during 8 min. of initial practice with work-loading as the parameter

.1636). Since the difference was significant beyond the 1% level it appears that work-loading is an effective variable *as early as the first 5 sec. of initial practice.*

Additional inspection of Fig. 2 reveals that (*a*) the difference in performance between the two extreme work-loads is about the same at the end of 8 min. of practice as it is at 1 min., (*b*) the initial difference between the two extreme conditions is about equal in magnitude to the total amount of decrement for any one curve alone, and (*c*) for at least 4 min., performance for those combinations with the lesser work-loads (21, 42, and 84) declines rapidly while performance remains at a near constant level for the heaviest work-load.

First 30-sec. scores of both practice periods.—The performances during the first 30 sec. of both initial and final practice are presented in Fig. 3. No trend is, of course, evident in the 30-sec. mean scores obtained during initial practice when the durations of initial practice and rest are concerned, the variability being attributable to sampling fluctuations. The trend in the work-loading means for initial practice is, as noted previously, significant.

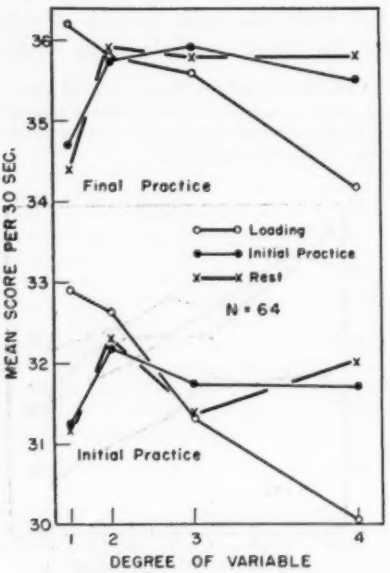

FIG. 3. Mean rates of responding as functions of the three experimental variables for the first 30 sec. of initial and final practice

TABLE 2

SPONTANEOUS RECOVERY AS A FUNCTION OF THREE EXPERIMENTAL VARIABLES

Each entry ($N = 64$) represents the mean difference between the first 30-sec. score of final practice and the last 30-sec. score of initial practice.

Variable	Degree of Variable			
	1	2	3	4
Initial practice	3.42	4.54	5.57	5.43
Interpolated rest	3.46	4.92	5.59	5.00
Work-loading	4.04	4.96	5.04	4.92

When the first 30-sec. scores of final practice are compared with those of initial practice it is immediately apparent that there was a large increase in mean score from initial to final practice. The highest scores obtained during the entire experiment were found in the initial portions of the final practice period, and these maximum scores occurred in final practice in spite of general decreases in rate of responding during initial practice.

A mean of approximately 32 balls was lifted during the first minute of initial practice and by the fourth minute a decrement of less than one ball had been incurred. In contrast the mean of the first minute of final practice was well over 34 balls per 30 sec. and a subsequent drop of approximately three balls had occurred by the fourth minute. It was apparent also that the rate of respond-

ing during the fourth minute of final practice was of the same order of magnitude as the highest scores obtained during initial practice (min. 1). Most noteworthy was the suggestion that the higher the initial level of responding the more rapid the decline in rate of responding.

Final practice: Spontaneous recovery.—Measures of spontaneous recovery for each of the three variables were obtained by subtracting the terminal 30-sec. score of initial practice from the first 30-sec. score of final practice. The recovery functions, presented in Table 2, are all initially increasing with a small decline corresponding to the largest value of initial practice, rest or work-loading.

Final practice: Warm-up decrement.—Inspection of the 48 curves (the 48 oblongs of a $4 \times 4 \times 4$ factorial design) for final practice revealed no evidence of the initial warm-up decrement which is often characteristic of pursuit rotor performance curves obtained under similar experimental conditions. Since these curves were plotted with respect to 30-sec. scoring intervals, it was conceivable that a warm-up decrement was obscured by the use of this measure. But warm-up was not indicated when the first 30-sec. period of final practice was fractionated into six 5-sec. scoring periods for the combination 2-8-*K* and for the combination 16-8-*K*. The latter combination is most representative of conditions favorable for the need to warm up, according to a view developed by Ammons (1), for it represents the longest period of practice prior to rest, and the longest rest prior to resumption of practice. The mean number of balls lifted per successive 5-sec. periods for 2-8-*K* was 6.3, 5.8, 5.6, 5.9, 5.9, and 5.7, while the corresponding means for 16-8-*K* were 6.1, 5.9, 6.1, 5.9, 5.8, and 5.6. These means, which should progressively increase in magnitude were the need to warm up in effect, actually show a tendency to decline in magnitude.

Other final practice results.—The effects of duration of initial practice, duration of interpolated rest and work-loading during final practice can be independently analyzed by summation over two of the three orthogonal variables.

1. *Duration of initial practice.*—Inspection of the final practice curves for 2-*K*-*K*, 4-*K*-*K*, 8-*K*-*K*, and 16-*K*-*K* revealed so little systematic difference among the curves that a test for significance of the differences was judged unnecessary.

2. *Duration of interpolated rest.*—Rate of responding throughout final practice with rest as the parameter is presented in Fig. 4. For at least 4 min., it appears that level of responding is an increasing function of the length of the rest period although there are inversions between rest of 2

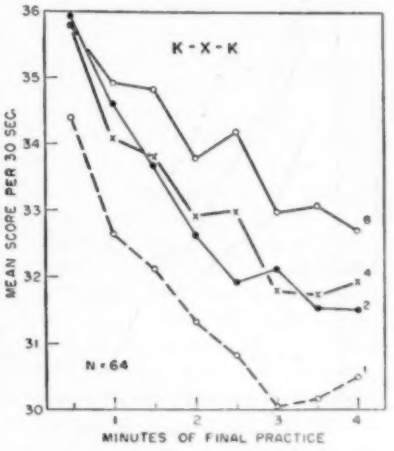

FIG. 4. Mean rate of responding during final practice with duration of interpolated rest as the parameter

and 4 min. Obviously, the effects of the rest period are not transitory but persist throughout the final test period.

3. *Work-loading.*—The effects of ball weight are illustrated in Fig. 5. The inverse relation between weight and performance noted in Fig. 2 for initial practice is again suggested, particularly when the most divergent weights are considered. The data of Fig. 5 also suggest that the difference between groups remains fairly constant throughout final practice, a finding not uncommon with duration of the work period as the variable in distribution of practice studies.

When the curves of Fig. 5 are compared with those of Fig. 2, two different, but perhaps not unrelated, observations can be made. The initial score of final practice for each load is higher than the corresponding load score for initial practice and the curves are also all more steeply pitched.

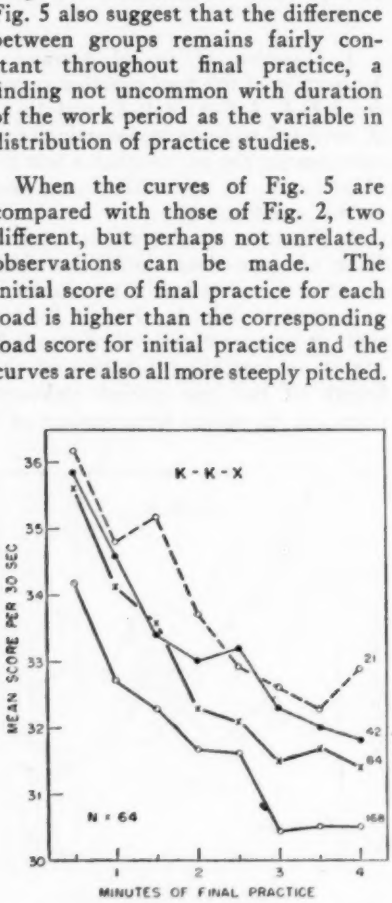

FIG. 5. Mean rate of responding during final practice with work-loading as the parameter

The first 30-sec. score of final practice was larger than the first 30-sec. score of initial practice for 89% of the *S*s, equal for 4%, and smaller for 7%. Separate analyses of the 89% and 7% categories were made to test the dependence of amount of decrement upon starting rate. Of those *S*s increasing their rate after rest, 77% had a greater post- than prerest 2-min. decrement; whereas, of those having a greater starting rate for initial practice, 78% had a greater prerest 2-min. decrement. Evidently the slope of the performance curve is dependent upon the magnitude of the first score of the initial or final practice session.

Although the trend in Fig. 5 is generally negative, a great many *S*s manifested a relatively large *but temporary* increase in score after 1 to 2 min. of final practice. No statistical test of this upward oscillation seemed appropriate, however. When the scores for each trial were summed over one of the experimental variables, 47 of the 48 combinations seemed to yield an indication of the short upward trend. If the trend is assumed to be real for a moment, then the entire postrest practice curve can be described by the *converse* of Ammons' (1, p. 263) description of an idealized postrest *pursuit rotor* (learning) curve, i.e., (*a*) an abrupt descent much faster than that to be expected if there had been no rest, (*b*) a short period of improvement in performance, and (*c*) resumption of the prerest curve of decreasing performance or at least some signs of this. Some additional evidence on the validity of the short upward trend has been presented elsewhere, but little confidence can be placed in this suggested phenomenon without further confirmation (3).

Discussion

Of the findings presented in the previous section only a few need to be considered for an extended discussion.

Hull's conception of the properties of I_R and ${}_sI_R$ leads to the prediction that rate of responding should be inversely related to ball weight during the early portions of initial practice.[5] The prediction is based upon the observation that a number of weights are lifted during *S*'s lifetime, and upon the assumptions concerning I_R and ${}_sI_R$. Specifically, it can be said that a certain amount of I_R is incurred each time an individual makes a response. Between responses, I_R has an opportunity to dissipate. But with the dissipation of I_R, the resting response (${}_sI_R$) is strengthened. One can also assume that objects of different weights are often lifted, and that under certain circumstances more I_R is generated the heavier the object manipulated. If, as supposed by Kimble (6), the quantity of ${}_sI_R$ at any moment is proportional to the amount of I_R which has been dissipated, then more ${}_sI_R$ will accrue to heavier weights. As *S* enters the ball-lifting situation and lifts the first ball, the ${}_sI_R$ is activated according to the similarity of the cues common to the ball weight and previous experience with this weight. In other words, the lifting of a ball will call out the resting response in proportion to the amount of ${}_sI_R$ previously conditioned to a similar weight. It is expected that the greater the difference between weights, the more dissimilar the activation of ${}_sI_R$, and consequently, the greater the initial difference in rate of responding for *S*s in conditions with such disparate weights. This, then, becomes one way of accounting for the rapidly developed inverse relation between rate of responding and work-load.

There are additional and perhaps better and simpler ways of accounting for the rapidly developed load difference. One is to attribute the inverse relationship to the effects of a *very* rapidly developed I_R. Another way is to suppose that in spite of the common form of the balls the load difference leads to differences in manipulability. Thus, attempts at speedy grasping, lifting, and dropping of a relatively small but heavy load might be expected to lead to increases in partial fumbles both near the cup and at the chute top. Such a phenomenon would, of course, be reflected by the rate measure employed, yet would not be directly measured.

The data of both initial and final practice favor an interpretation of little or no interaction between work-load and practice time. Such a finding is probably surprising to those who may wish to invoke ${}_sI_R$ in the explanation of these data. The most that can be said about the intercomparison of slopes is that there appears no great difference in total inhibition from load group to load group as a function of practice time. Recently completed research by the writer has verified this finding and it will be discussed in a subsequent paper.

After interpolated rest not only did *S*s of all conditions regain their initial rates of responding but actually far surpassed their previous high score. The increase of the first 30-sec. scores of final practice over the previous high scores of initial practice might be interpreted as due to an increase in ${}_sH_R$ during initial practice or to an increase in D during rest. If the increase in performance is due to ${}_sH_R$ one

[5] The intervening variables I_R, ${}_sI_R$, ${}_sH_R$, D, and ${}_s\bar{E}_R$ are defined here after Hull (4) and are used after Kimble (6).

must assume that learning was indeed taking place during initial practice even though the strength of the response was decreasing progressively throughout the practice session. The increase in ${}_{S}H_{R}$ must then have occurred rapidly inasmuch as *Ss* who practiced for only 4 min. reached the same rate of responding during final practice as those practicing 8 or 16 min. If the sizable increase of the first 30-sec. score of final practice is interpreted as a change in motivational level brought about by the interpolated rest one might speculate that it was at least partly dependent on one feature of the experimental procedure; namely, that during the interpolated rest period *Ss'* counters were reset to zero. The suggestion here is that this feature remotivates an *S* to out-distance his partner.

It was shown that rate of responding decreased rapidly for about the first minute of final practice. This observation is discussed below with respect to two somewhat different topics:

1. *Growth of inhibition.*—During final practice decreases in rate of responding were generally greater than for initial practice while the first 30-sec. scores were higher. This can be interpreted as evidence that the growth of I_R with continuous practice is some increasing function of the level of ${}_{S}\bar{E}_{R}$. It is still conceivable, however, that learning may have masked a portion of the decremental effects of initial practice.

2. *Warm-up decrement.*—The results have shown little evidence for warm-up decrement, at least in terms of a common definition of such an event, i.e., a rapid postrest increase in score from Trial 1 to Trial 2. A number of somewhat different leads are available for further study of the warm-up decrement phenomenon.

Ammons (1, p. 271) proposes to account for initial decrement during final practice by supposing that "warm-up decrement is owing to loss during rest of 'set,' consisting principally of advantageous postural adjustments." He cites evidence (2) in support of the notion that these sets are acquired during initial practice and are dissipated during rest. In his review of the pursuit-rotor literature he also presents evidence that well-practiced *Ss* show little if any warm-up decrement (9).

Lewis and Shephard (7), using the Mashburn apparatus, have investigated initial decrements after rest by attempting to manipulate *Ss'* activity during the rest (interpolated practice) period. During the interpolated period control-indicator relationships were reversed; later testing used the original arrangement of the controls. The rate of rise of the matches curve, for example, was much faster after interpolated learning than before, but the number of matches was initially much lower than during the terminal trials of original learning. This experiment suggests the interference of one set of habits upon another when the interfering habits are introduced by *E*. It is possible, then, that *S* can enter a motor learning situation with a previously acquired set of interfering habits which serves to depress performance to a level lower than would otherwise obtain were the interference absent. These habits may be partially or wholly extinguished during initial practice, but may spontaneously recover during rest and as a consequence depress performance during the early trials of postrest as well as initial practice. Conceivably, well-practiced tasks performed under relatively high degrees of motivation should produce little or no warm-up decrement inasmuch as the response tendency is at or near maximum strength and competing habits are relatively low in the response hierarchy.

Still another attack upon the nature of warm-up decrement is conceptually possible: *S* is allowed in the ball-lifting task to pace himself and is not required to set up motor rhythms coordinate with a revolving disk. Irion (5), for example, has presented some evidence that less warm-up decrement is obtained when learning paired associates *if* rhythmical color naming is required during the interpolated rest period. This may indicate that the breaking of a rhythmical pattern (by rest) may at least be one of several factors operating to produce the observed decrement. In this connection it may be noted that Kimble (6) obtained no evidence of warm-up decrement in an unpaced task (inverted alphabet printing).

It has now been indicated that a number of factors acting singly or in combination may currently be used in accounting for warm-up decrement. In summary, these factors include the loss of set, interference, and the breaking of pace. The listing of these factors, however, is in no way intended to indicate that a general principle cannot be formulated which includes the more serviceable parts of each (3).

Summary

A total of 256 women practiced a relatively simple motor task in which *Ss* raced in lifting weighted table-tennis balls from a cup at the base of a chute to the top of the chute. Three experimental variables were manipulated in a factorial design: (*a*) duration of a continuous initial practice session (2, 4, 8, or 16 min.), (*b*) duration of a single interpolated rest period (1, 2, 4, or 8 min.) and (*c*) the weight (work-load) of the lead-filled table-tennis balls (21, 42, 84, or 168 gm.). A final practice period of 4 min. completed the session.

During initial practice the scores decreased progressively for about 8 min. (for the 8- and 16-min. practice conditions) and subsequently appeared to remain fairly constant (for the 16-min. practice conditions). The first 30-sec. scores of final practice were the highest scores obtained during the entire session for any combination of conditions. During the ensuing practice the scores decreased rapidly and regularly in magnitude except perhaps for a short period of improvement.

The three experimental variables appeared to have the following effects upon performance:

1. *Duration of the initial practice period.*—Little or no difference in mean scores during final practice could be attributed to the length of the initial practice period. A measure of spontaneous recovery appeared to be an increasing function of the length of this variable except for a small decline for the longest practice period.

2. *Duration of the interpolated rest period.*—Spontaneous recovery from the decremental effects of initial practice appeared to be an increasing function of the length of the interpolated rest period except for a small decline for the longest rest period. Evidence was presented that the scores of the entire final period were directly related to the duration of the interpolated rest.

3. *Work-loading.*—The scores of both practice periods were generally inversely related to the ball weight. The inverse relation was evidenced clearly in the 30-sec. scores of the initial practice period. Further, a significant difference between loads 21 and 168 was obtained for the first 5-sec. period of initial practice. Decrements after rest were greater than before rest, possibly because rate of responding was initially faster after rest. The course of spontaneous recovery was similar to that noted above for the other two variables.

A discussion of the more critical results was presented, based on certain of Hull's concepts, which it is believed is in substantial agreement with findings obtained with the pursuit rotor as well as with the present task. The roles of I_R, ${}_sI_R$ and competing habits were stressed in their relationship to ${}_sE_R$ and to ${}_s\bar{E}_R$ under conditions where the level of ${}_sH_R$ can be presumed to be high and fairly constant.

(Manuscript received August 23, 1951)

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**(1), No. 243.
3. Bilodeau, E. A. Performance in a simple motor task before and after a single rest period. Unpublished Ph.D. thesis. State Univer. Iowa, 1950.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

5. Irion, A. L. Retention and warming-up effects in paired associate learning. *J. exp. Psychol.*, 1949, **39**, 669–675.
6. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
7. Lewis, D., & Shephard, A. H. Devices for studying associative interference in psychomotor performance: I. The modified Mashburn apparatus. *J. Psychol.*, 1950, **29**, 35–46.
8. Robinson, E. S. Work of the integrated organism. In C. Murchison (Ed.), *A handbook of general experimental psychology*. Worcester, Mass.: Clark Univer. Press, 1934.
9. Travis, R. C. Practice and rest periods in motor learning. *J. Psychol.*, 1937, **3**, 183–187.

TRANSFER OF WORK INHIBITION IN MOTOR LEARNING

GREGORY A. KIMBLE

Yale University

This experiment was performed to test the hypothesis that the inhibition generated in motor learning under conditions of massed practice is general rather than specific in that it influences the performance of effectors other than those actually exercised in its development. This assumption, which is to the effect that work decrement is not entirely fatigue in the usual sense of *localized muscle fatigue*, was suggested by an article of Cason (1). If the hypothesis is correct, it means that responses other than those just performed will be depressed by the accumulation of inhibition, i.e., that the decrement developed by one responding mechanism will transfer to others.

A preliminary test of this hypothesis seems to be provided in the case of skill acquired through the mechanism of bilateral transfer where a motor response is learned without overt participation of the muscles used to perform it. If it can be shown that the decrement developed in a practiced hand transfers bilaterally, this is evidence that such decrements are more general than a local fatigue theory would lead one to suppose, and would argue against the notion that work decrement is confined to practiced effectors. The prediction to which this theory of nonlocalized work decrement leads is that, immediately after practice with one hand, the skill acquired in the other by transfer will be depressed; further that such skill acquired by transfer will show "reminiscence" with rest. The present experiment is a test of these predictions for rotary pursuit performance.

Method

A modified transfer of training method involving two groups of 15 *Ss* each was used. All *Ss* were given 60 10-sec. trials on the rotary pursuit apparatus; the 60 trials being divided into two blocks of 30 massed trials each. Group I practiced 30 trials on the rotor with the nonpreferred hand, rested for 5 min., and then practiced for 30 trials with the preferred hand. Group II practiced for 30 trials with the nonpreferred hand and then for 30 trials with the preferred hand. But for this group no rest pause in the usual sense was allowed. The transfer of the stylus from hand to hand was accomplished in 10 sec. The *Ss* were volunteers from the course in elementary psychology at Yale University.

The apparatus was a typical modification of the Koerth pursuit rotor. Continuous records of performance were made possible by the use of two clocks and a cycling mechanism such that *S* cumulated a score for 10 sec. on one clock while the other was being read and reset. Practice was thus continuous for all groups, except where a rest was introduced deliberately or made necessary for purposes of changing the practice conditions.

Results and Conclusions

Performance curves for the two groups of *Ss* are plotted in Fig. 1. In the left-hand portion of the graph, the data for Groups I and II have been combined into a single function because the practice conditions were the same and because the two sets of data were very similar, beginning at exactly the same mean value and showing nearly identical improvement trends. In the right-hand section of the figure, performance curves are plotted separately. In this section of the graph there is an obvious difference between the curves for Groups I and II. Beginning with Trial 31, Group I is clearly superior to Group II. On this trial, the mean difference is 1.1 sec. This difference yields a *t*-value of 3.72 which is significant at the .001 level. On Trial 60, the difference still approaches significance. The mean difference is .7 sec. which is reliable between the .1 and .2 levels of confidence.

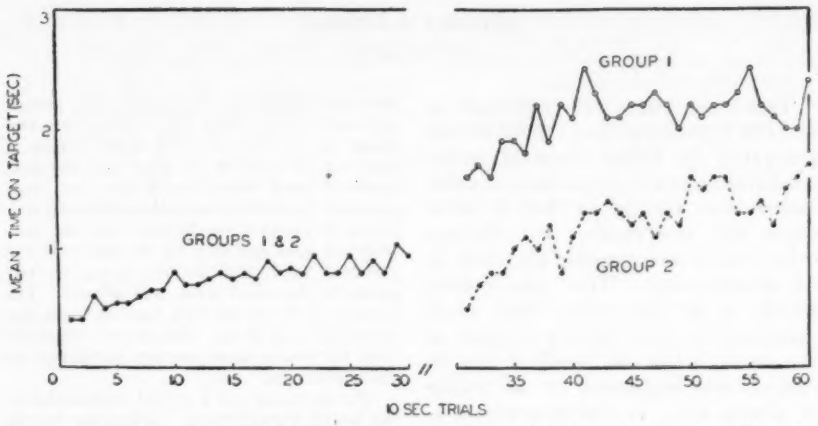

FIG. 1. Mean time on target (*a*) during learning with the left hand (both groups), and (*b*) during test of Group 1 after a rest and Group 2 after no rest when practice is with the right hand

These results are consistent with the hypothesis that the typical decremental effect of massed practice is general in that it transfers to muscle groups not actually exercised in the build up of decrement. The conclusion that the depression of the motor habit acquired by transfer is a function of the same decremental factor as is commonly encountered in motor learning is based upon the facts that: (*a*) this decrement is generated by massed practice, and (*b*) it dissipates with rest. The conclusion that the decrement is general depends upon the assumption that the nonpracticed hand is relatively inactive during practice with the other hand.

The most important question raised by this experiment is the following: Is this transfer of decrement a phenomenon associated with transfer of learning, or will it also occur in more highly practiced skills involving responding mechanisms anatomically less closely related? In brief, just how general is this generalized work decrement?

SUMMARY

Two groups of 15 male *Ss* were run under two different conditions on the pursuit rotor. Each group received 60 10-sec. trials. For Group I, half of these were massed trials with the nonpreferred hand, and half were massed trials with the preferred hand after a 5-min. rest. For Group II, the practice arrangement was the same except that no rest between the two groups of trials was allowed. From the beginning of the second practice session, Group II was significantly inferior to Group I. This fact suggests that the decremental effect of massed practice transfers to response systems other than those specifically exercised in the generation of the decrement. This generality argues against the position that the decrements known to result from massed practice in motor learning are localized muscle fatigue and nothing else.

(Manuscript received September 18, 1951)

REFERENCE

1. CASON, H. The organic nature of fatigue. *Amer. J. Psychol.*, 1935, **47**, 337–342.

New McGRAW-HILL *Books*

PSYCHOLOGY

By Ross Stagner, University of Illinois, and Theodore Karwoski, Dartmouth College. 582 pages, $5.00

An outstanding new text for first-year college students, *Psychology* covers the facts of psychology in its important areas (motivation, sensing, perceiving, learning, thinking, intelligence, and personality) and presents these facts in such a way that the student can readily grasp both their interrelationships and significance within the total scope of the book itself. In order to effect the exceptional unity of treatment which this book affords, the authors have presented their material in terms of *homeostasis* and the *psychology of perception.*

STATISTICAL THEORY IN RESEARCH

By R. L. Anderson, University of North Carolina, and T. A. Bancroft, Iowa State College. In press

Divided into two parts, this text first presents basic statistical theory, including the elementary principles of probability, population and sampling distribution theory with moment-generating functions, orthogonal linear forms, and the theory of estimation and tests of significance. A chapter on the theory behind the uses of Chi-square is included. Part II presents the theory of least squares and its use in the analysis of actual experimental data, including multiple regression, experimental design and variance component models.

THE MEASUREMENT OF HEARING

By Ira J. Hirsh, Washington University Medical School. *McGraw-Hill Publications in Psychology.* Ready in July

A text for clinical otologists, experimental psychologists, acoustics and communications engineers, and others who do experimental work on hearing, this text brings together basic, experimental information about acoustics, electro-acoustic equipment, psychology of hearing, etc. and applies this information to the several facets of the measurement of hearing. Offers basic information from physics and psychology pertinent to the student's clinical problems.

THE PSYCHOLOGY OF LEARNING

By James E. Deese, Johns Hopkins University. *McGraw-Hill Publications in Psychology.* Ready in September

A broad treatment of the psychology of learning, written from the point of view of an experimental psychologist and covering both human and animal learning. Emphasis is on experimental evidence rather than upon current theoretical "schools." Recent experimental literature is stressed, and post war experimental work on such topics as punishment, retroactive inhibition, etc. is highlighted.

Send for copies on approval

McGRAW-HILL BOOK COMPANY, Inc.

330 West 42nd Street **New York 36, N.Y.**